500.43155X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): N. Watanabe

Serial No.: 10/669,325

Filed: September 25, 2003

Title: REMOTE COPY SYSTEM

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

November 6, 2003

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby claim(s) the right of priority based on:

Japanese Patent Application No. 2003-205617
Filed: August 4, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Carl I. Brundidge
Registration No.: 29,621

CIB/rr
Attachment

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: ２００３年　８月　４日

出願番号
Application Number: 特願２００３－２０５６１７

[ST. 10/C]: ［ＪＰ２００３－２０５６１７］

出願人
Applicant(s): 株式会社日立製作所

２００３年１０月　２日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号　出証特２００３－３０８１２０６

【書類名】　特許願
【整理番号】　K03008941A
【あて先】　特許庁長官殿
【国際特許分類】　G06F 12/00
【発明者】
　【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所システム開発研究所内
　【氏名】　渡辺　直企
【特許出願人】
　【識別番号】　000005108
　【氏名又は名称】　株式会社　日立製作所
【代理人】
　【識別番号】　100075096
　【弁理士】
　【氏名又は名称】　作田　康夫
【手数料の表示】
　【予納台帳番号】　013088
　【納付金額】　21,000円
【提出物件の目録】
　【物件名】　明細書　1
　【物件名】　図面　1
　【物件名】　要約書　1
【プルーフの要否】　要

【書類名】明細書

【発明の名称】リモートコピーシステム

【特許請求の範囲】

【請求項１】

計算機に接続され、第1のディスク装置と第1のコントローラとを有する第1の記憶装置システムと、

第2のディスク装置と第2のコントローラとを有する第2の記憶装置システムと、

前記第1の記憶装置システム及び前記第2の記憶装置システムとに接続される、第3の記憶領域と第3のコントローラとを有する第3の記憶装置システムとを有するシステムであって、

前記第1のコントローラは、

前記計算機から受信したライト要求に応じて、前記計算機から受信したライトデータと該ライトデータが書き込まれる位置を示すアドレス情報とを含むジャーナルを前記第3の記憶装置システムに送信し、

前記ライトデータを前記第1のディスク装置に格納し、

前記ジャーナルの送信後に前記計算機に前記ライト要求に対する応答を返し、

前記第2のコントローラは、

前記第1のコントローラによって発行され、前記ジャーナルを前記第2の記憶装置システムが取得する際に使用される、前記ジャーナルの格納位置を含む第1の制御情報を受信し、

該第1の制御情報に基づいて、前記ジャーナルを前記第3の記憶サブシステムから取得し、

前記ジャーナルに含まれるアドレス情報に基づいて前記ライトデータを前記第2のディスク装置に格納することを特徴とするシステム。

【請求項２】

請求項１記載のシステムにおいて、

前記第1の記憶装置システムは、前記第1の制御情報を格納しており、

前記第2のコントローラは、前記ジャーナルを取得した後、その旨を示す第2の

制御情報発行し、

前記第1のコントローラは、前記第2の制御情報を取得した後、前記第1の記憶装置システム内に格納されている前記第1の制御情報を廃棄可能とすることを特徴とするシステム。

【請求項３】

請求項2記載のシステムにおいて、

前記第1のコントローラは前記第3の記憶装置システムに前記第1の制御情報を送信し、前記第2のコントローラは前記第3の記憶領域に格納された前記第1の制御情報を前記第3の記憶装置システムから取得し、

前記第2のコントローラは前記第2の制御情報を前記第3の記憶装置システムに送信し、前記第1のコントローラは前記第3の記憶領域に格納された前記第2の制御情報を前記第3の記憶装置システムから取得することを特徴とするシステム。

【請求項４】

請求項３記載のシステムにおいて、

前記第3の記憶装置システムは、前記第1の制御情報及び前記第2の制御情報を前記第3の記憶領域内の各々異なる論理ボリュームに格納し、論理ボリュームごとに前記第1の記憶装置システムもしくは前記第2の記憶装置システムのいずれからの書き込み要求を許可するかを設定することを特徴とするシステム。

【請求項５】

請求項３記載のシステムにおいて、

前記第1の記憶装置システムに障害が生じた場合に、前記第2の記憶装置システムは、前記第3の記憶装置システムに格納されている第1の制御情報を参照して、前記第2のディスク装置に格納されていないライトデータを有するジャーナルを前記第３の記憶装置システムから取得し、取得したジャーナルに含まれるアドレス情報に基づいて、取得したジャーナルに含まれるライトデータを前記第2のディスク装置に格納することを特徴とするシステム。

【請求項６】

請求項５記載のシステムにおいて、

前記第2の記憶装置システムは、前記第1の記憶装置システムに障害が生じた後

に、前記第2の記憶装置システムに接続される計算機からライト要求を受信した場合に、該ライト要求に従って書き込まれるライトデータの格納位置を示す差分情報を有しており、

前記第1の記憶装置システムが障害から回復した場合に、前記第２の記憶装置システムは、前記差分情報が示す格納位置に格納されているデータを、前記第1の記憶装置システムと前記第2の記憶装置システムとを接続する通信路を介して、前記前記第1の記憶装置システムに送信することを特徴とするシステム。

【請求項７】

請求項2記載のシステムにおいて、

更に前記第1の記憶装置システムと前記第2の記憶装置システムとに接続される通信路を有し、

前記第1のコントローラは前記第1の制御情報を前記通信路を介して前記第2の記憶装置システムに送信し、

前記第2のコントローラは前記第2の制御情報を前記通信路を介して前記第1の記憶装置システムに送信することを特徴とするシステム。

【請求項８】

請求項７記載のシステムにおいて、

前記第３の記憶装置システムに障害が生じた場合に、前記第１のコントローラは計算機から受信するライトデータを前記通信路を介して前記第２のコントローラに送信し、前記第２のコントローラは前記通信路を介して受信したライトデータを前記第２のディスク装置に格納することを特徴とするシステム。

【請求項９】

請求項3記載のシステムにおいて、

更に前記第1の記憶装置システムと前記第2の記憶装置システムとに接続される第４の記憶装置システムを有し、

前記第３の記憶装置システムに障害が発生した場合に、前記第１の記憶装置システムと前記第２の記憶装置システムは、前記第４の記憶装置システムを介して、ジャーナル、第１の制御情報、及び第２の制御情報を送受信することを特徴とするシステム。

【請求項１０】

請求項３記載のシステムにおいて、

更に前記第１の記憶装置システムと前記第２の記憶装置システムとに接続される第４の記憶装置システムを有し、

前記第１のコントローラは時刻情報を有するジャーナルを前記第３の記憶装置システム若しくは前記第４の記憶装置システムのいずれかに送信し、

前記第２のコントローラはジャーナルを前記第３の記憶装置システム及び前記第４の記憶装置システムから取得して、取得したジャーナルに含まれるライトデータを、該ジャーナルに付与された時刻情報が示す時刻順に前記第２のディスク装置に書き込むことを特徴とするシステム。

【請求項１１】

計算機に接続され第1のディスク装置を有する第1の記憶装置システムと、第2のディスク装置を有する第2の記憶装置システムと、前記第1の記憶装置システム及び前記第2の記憶装置システムとに接続される第3の記憶装置システムとを有するシステムにおいて、前記第１の記憶装置システム、前記第２の記憶装置システム、及び前記第３の記憶装置システム間で実行されるリモートコピー方法であって、

前記第１の記憶装置システムが前記計算機からライト要求とライトデータとを受信するライト要求受信ステップと、

前記第１の記憶装置システムが、前記ライトデータと前記ライト要求に含まれるアドレス情報とを有するジャーナルを前記第３の記憶装置システムに書き込むジャーナル書き込みステップと、

前記ジャーナルを前記第２の記憶装置システムが読み出すために必要な、前記ジャーナルの格納位置を含む第１の制御情報を、前記第１の記憶装置システムが発行する第１制御情報発行ステップと、

前記第２の記憶装置システムが、前記第１の制御情報を取得する、第１制御情報取得ステップと、

前記第２の記憶装置システムが、前記第１の制御情報に基づいて、前記ジャーナルを読み出すジャーナル読み出しステップと、

前記第２の記憶装置システムが、前記ジャーナルに含まれるアドレス情報に従って、前記ジャーナルに含まれるライトデータを該第２の記憶装置システムが有するディスク装置に格納するライトデータ書き込みステップとを有することを特徴とするリモートコピー方法。

【請求項１２】

請求項１１記載のリモートコピー方法において、

前記第1の記憶装置システムは、前記第1の制御情報を格納しており、

更に、前記ジャーナル読み出しステップ後、前記第2の記憶装置システムが該ジャーナルを読み出した旨を示す第2の制御情報発行する第２制御情報発行ステップと、

前記第1の記憶装置システムが、前記第２の制御情報を受信する第２制御情報受信ステップとを有し、

前記第１の記憶装置システムは前記第２の制御情報を受信した後に前記第１の制御情報を破棄することを特徴とするリモートコピー方法。

【請求項１３】

請求項１１記載のリモートコピー方法において、

前記第１制御情報発行ステップは、前記第１の制御情報を前記第３の記憶装置システムに書き込むステップを有し、

前記第１制御情報取得ステップは、前記第３の記憶装置システムから前記第１の制御情報を読み出すステップを有することを特徴とするリモートコピー方法。

【請求項１４】

請求項１３記載のリモートコピー方法において、

前記第3の記憶装置システムは、前記第1の制御情報及び前記第2の制御情報を前記第3の記憶領域内の各々異なる論理ボリュームに格納し、論理ボリュームごとに前記第1の記憶装置システムもしくは前記第2の記憶装置システムのいずれからの書き込み要求を許可するかを設定することを特徴とするリモートコピー方法。

【請求項１５】

請求項１３記載のリモートコピー方法において、更に、

前記第1の記憶装置システムに障害が生じた場合に、前記第2の記憶装置システムが、前記第3の記憶装置システムに格納されている第1の制御情報を参照して、該第2の記憶装置システムが有するディスク装置に格納されていないライトデータを有するジャーナルを前記第３の記憶装置システムから取得するステップと、

取得したジャーナルに含まれるアドレス情報に基づいて、取得したジャーナルに含まれるライトデータを前記第2の記憶装置システムが有するディスク装置に格納するステップとを有することを特徴とするリモートコピー方法。

【請求項１６】

請求項１５記載のリモートコピー方法において、

前記第1の記憶装置システムに障害が生じた後、前記第２の記憶装置システムが前記第2の記憶装置システムに接続される計算機からライト要求を受信するステップと、

前記第２の記憶装置システムが、前記ライト要求に従って書き込まれるライトデータの格納位置を示す差分情報保持するステップと、

前記第1の記憶装置システムが障害から回復した場合に、前記差分情報が示す格納位置に格納されているデータを、前記第1の記憶装置システムと前記第2の記憶装置システムとを接続する通信路を介して、前記前記第1の記憶装置システムに送信するステップとを有することを特徴とするリモートコピー方法。

【請求項１７】

請求項１１記載のリモートコピー方法において、

前記第1の記憶装置システムと前記第2の記憶装置システムとは通信路によって接続されており、

前記第１制御情報発行ステップは、前記第1の記憶装置システムが前記第1の制御情報を前記通信路を介して前記第2の記憶装置システムに送信するステップを有し、

前記第１制御情報取得ステップは、前記第２の記憶装置システムが前記第１の制御情報を前記通信路を介して前記第１の記憶装置システムから受信するステップを有することを特徴とするリモートコピー方法。

【請求項１８】

請求項１７記載のリモートコピー方法において、前記第３の記憶装置システムに障害が生じた場合に更に、

前記第１の記憶装置システムは前記計算機から受信するライトデータを前記通信路を経由して前記第２の記憶装置システムに送信するステップと、

前記第２の記憶装置システムが前記通信路を経由して受信したライトデータを該第２の記憶装置システムが有するディスク装置に格納するステップとを有することを特徴とするリモートコピー方法。

【請求項１９】

請求項１３記載のリモートコピー方法において、

前記システムは更に、前記第1の記憶装置システムと前記第2の記憶装置システムとに接続される第４の記憶装置システムを有し、

前記第３の記憶装置システムに障害が発生した場合に、前記第１の記憶装置システムと前記第２の記憶装置システムは、前記第４の記憶装置システムを介して、ジャーナル及び第１の制御情報を送受信することを特徴とするリモートコピー方法。

【請求項２０】

請求項１３記載のリモートコピー方法において、

前記システムは更に前記第１の記憶装置システムと前記第２の記憶装置システムとに接続される第４の記憶装置システムを有し、

ジャーナル書き込みステップは、前記第１の記憶装置システムが時刻情報を有するジャーナルを前記第３の記憶装置システム若しくは前記第４の記憶装置システムのいずれかに書き込むステップを有し、

前記第２のコントローラはジャーナルを前記第３の記憶装置システム及び前記第４の記憶装置システムから取得して、取得したジャーナルに含まれるライトデータを、該ジャーナルに付与された時刻情報が示す時刻順に前記第２のディスク装置に書き込むことを特徴とするリモートコピー方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明はコンピュータによって使用されるデータを格納する記憶装置に関する。特に、コンピュータとの間でデータの授受を行う制御部と、データを格納するディスク装置とを内蔵する記憶サブシステムと、この記憶サブシステムから距離的に離れた場所に存在する他の記憶サブシステムとを相互に接続し、コンピュータを経由せずに記憶サブシステム間でデータをコピーして、データを二重化するリモートコピー技術に関する。

【０００２】

【従来の技術】

大型計算機システム、サーバー、ネットワーク上のパーソナルコンピュータ、又はその他の上位計算機システム（以下、ホストという。）からデータを受け取った記憶装置（以下、記憶サブシステムという。）が、遠隔地に設置された第２の記憶サブシステムに対して、データの順序性を保証しながら受信したデータを非同期転送し、第２の記憶サブシステムが転送されたデータを書き込む、非同期リモートコピーの技術がある。ここで非同期転送とは、記憶サブシステムが、ホストからデータを受信し、この受信データに対する処理の完了を通知する応答をホストに返信した後に、第2記憶サブシステムへのデータ転送を実行することを意味する。

【０００３】

また、ホストとこれに接続された第１記憶サブシステム間のデータ更新処理と同期させて、第１記憶サブシステムと第１記憶装置システムの付近地又は遠隔地に設置された第2の記憶サブシステムとの間でデータを同期転送する同期リモートコピー技術もある。ここで同期転送とは、記憶サブシステムが、ホストからデータを受信し、受信データを第2記憶サブシステムに転送した後に、ホストに対して応答を返信することを意味する。同期リモートコピー技術を用いれば、巨視的にみて２つの記憶サブシステムに格納されているデータが一致しており、データの書込み順序も保証されている。尚、適当なデータ転送経路を選択すれば、２つの記憶サブシステムの距離が１００ｋｍを超える場合であっても、同期転送によるコピーが可能である。

【０００４】

更に、3以上の記憶サブシステム間で、同期リモートコピーと非同期リモートコピーを組み合わせて、データ更新の順序性を保証したデータの二重化を実現する技術が特許文献１及び特許文献２に開示されている。

【０００５】

【特許文献１】

特開２０００-３０５８５６号公報

【特許文献２】

特開２００３-１２２５０９号公報

【０００６】

【発明が解決しようとする課題】

従来技術では、2箇所のデータ格納拠点（サイト）間でのリモートコピーを複数組み合わせてｎサイト間でリモートコピーを実行する。

【０００７】

従って、各サイトにデータのコピーを保持する必要があるため、各サイトはリモートコピー処理を実行しない場合と比べて最低でもｎ倍の記憶容量を要し、コストが非常に高くなる。

【０００８】

またすべてのサイトにリモートコピープログラムを実装した記憶サブシステムが必要なため、高機能かつ高価な記憶サブシステムが複数台必要になる。

【０００９】

ｎサイトに渡る複数のリモートコピーペアの状態を監視・制御する必要があるため管理・制御が複雑となり開発コストも高くなる。

【００１０】

また、一般にリモートコピーに係わる処理の負荷は大きいため、正サイトと副サイトの中間に位置し、正サイトとも副サイトともリモートコピー処理を実行する必要のある中間サイトの負荷は特に大きくなり、中間サイトの記憶サブシステムが処理可能なI/O数が制限されてしまう。

【００１１】

従って、より適切なnサイト間でのリモートコピー技術が求められている。

【００１２】

【課題を解決するための手段】

第１の記憶装置システムと第２の記憶装置システムとを第３の記憶装置システムを介して接続する。リモートコピー処理を実行する際、第１の記憶装置システムは計算機から受信したライト要求に応じて、計算機から受信したライトデータとライトデータが書き込まれる格納位置を示すアドレス情報とを有するジャーナルを第３の記憶装置システムに送信し、第３の記憶装置にライトする。第２の記憶装置システムは、第１の記憶装置システムによって発行される制御情報を受信し、制御情報に基づいて第３の記憶装置からジャーナルをリードして取得する。そして第２の記憶装置システムは、ジャーナルに含まれるアドレス情報に従って、ジャーナルに含まれるライトデータを、第２の記憶装置システム内のディスクに書き込む。

【００１３】

【発明の実施の形態】

－第１の実施形態－

図１に、本発明の第１の実施形態にかかるデータ処理システム（以下、リモートコピーシステムともいう。）の構成例を示す。なお、以下の説明では、正／副／中間を符号のａ，ｂ，ｃにより区別する。又、正／副／中間を区別しなくても良いときには符号のａ，ｂ，ｃを省く場合もある。

【００１４】

データ処理システムは、正ホストコンピュータ(以下、正ホスト若しくは第１ホストと呼ぶ。)１０２ａ及び正記憶サブシステム１０４ａ（以下、第１記憶サブシステムと呼ぶ場合もある。）を有する正サイト（以下、第１サイトと呼ぶ場合もある。）１０１ａ、副ホスト（以下、第２ホストと呼ぶ場合もある。）１０２ｂ及び副記憶サブシステム（以下、第２記憶サブシステムと呼ぶ場合もある。）１０４ｂを有する副サイト（以下、第２サイトと呼ぶ場合もある。）１０１ｂ、中間記憶サブシステム（以下、第３記憶サブシステムと呼ぶ場合もある。）１０４ｃを有する中間サイト（以下、第３サイトと呼ぶ場合もある。）１０１ｃを有する。

【００１５】

本実施形態では、記憶サブシステム１０４としてディスクアレイ装置を用いる例を説明するが、記憶サブシステム１０４は特にディスクアレイ装置に限られることはない。また、各サイト１０１内のホスト１０２、記憶サブシステム１０４の数に関しても、一以上であれば特に制限はない。また、各ホスト１０２は、ワークステーション、マイクロコンピュータ又はメインフレームコンピュータ等である。

【００１６】

記憶サブシステム１０４は一つ以上のインタフェース１１０を持ちストレージエリアネットワーク（ＳＡＮ１０３）を経由しホスト１０２と接続される。ここでインタフェース１１０はＳＣＳＩ、ファイバチャネル、ＦＩＣＯＮ、又はＥＳＣＯＮ等の記憶装置向けインタフェースであり、インタフェース１１０の種類およびＳＡＮ１０３の構成に関しては特に制限はない。以下本実施形態の説明においてはインタフェース１１０としてファイバチャネルを用いた例を説明する。

【００１７】

異なるサイト１０１に存在するの記憶サブシステム１０４間も同様にインタフェース１１０（異なるサイト１０１に存在する記憶サブシステム１０４間の接続に用いられるインタフェースを以下、リモートコピーリンク１１６とも呼ぶ。）を介して接続される。

【００１８】

正記憶サブシステム１０４ａと中間記憶サブシステム１０４ｃは同じ部屋の中、または同じビルの中にあっても良いが、安全性のために（両装置が同時に同じ障害にあわないよう）距離をおいても良い。正記憶サブシステム１０４ａと中間記憶サブシステム１０４ｃの間では同期リモートコピーが実行され、データの更新は同期で行われるため、正記憶サブシステム１０４ａと中間記憶サブシステム１０４ｃの距離は短い方がよい。しかし、当該距離はシステムの立地条件等の環境により決定されるべきであり、特に制限はない。尚、通常通信の伝播遅延は１ｋｍあたり５ｕｓといわれているので、当該距離が１００ｋｍ程度までであれば、ホスト１０２ａへ著しい影響を与えることは少ないといわれている。

【００１９】

副記憶サブシステム１０４ｂは、安全性向上のために、正記憶サブシステム１０４ａや副記憶サブシステム１０４ｂとは距離の離れた遠隔地に設置することが好ましい。

【００２０】

尚記憶サブシステム１０４間の距離が長く、ファイバチャネルのデータ転送可能距離を越える場合には、ファイバチャネルに加え、エクステンダ装置を介してＡＴＭ、ＳＯＮＥＴ、ＤＷＤＭ、ＩＰなどの広域回線を経由して、記憶サブシステム１０４間を接続しても良い。

【００２１】

管理コンソール１１７は記憶サブシステム１０４を監視・管理するためのソフトウェアであり、ＣＰＵ２０２やメモリ２０５を実装したホスト１０２上に実装される。管理コンソール１１７は災害時のことを考え，第四サイト（管理センタ）１０１ｄ上に設置され、第四サイト内のホスト上で稼動する。

【００２２】

管理コンソール１１７は管理ネットワーク１１１を介して各記憶サブシステム１０４と接続され、記憶サブシステム１０４の状態の監視したり、データ処理システムの構成変更を実施したり、リモートコピー処理を制御したりする。一般に管理ネットワーク１１１はＩＰプロトコルを用いたネットワークにて構成されるが、他のプロトコルを用いたネットワークであってもかまわない。また，本実施例ではリモートコピーリンク１１６は管理ネットワーク１１１とは別のネットワークとしているが、管理ネットワークとしてＩＰネットワーク等を用い、更に管理ネットワーク１１１をリモートコピーリンク１１６として使用しても良い。

【００２３】

本実施形態においては、正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂがリモートコピーのペアを構成している。一つの正記憶サブシステム１０４ａ内の正ボリューム１０９と一つ以上の副記憶サブシステム１０４ｂ内の副ボリューム１１５がリモートコピーペアとして関連付される。

【００２４】

リモートコピーの制御は記憶サブシステム１０４上のプログラム１０５によって実現される。正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂのプログラム１０５は基本的に同等の機能を持つので、同一記憶サブシステム１０４内に正ボリュームと副ボリューム１１５が混在していても良いし、災害又はメンテナンス時には、正ボリューム１０９と副ボリューム１１５の切り替えを実行することもできる。

【００２５】

リモートコピーを実現するために，正記憶サブシステム１０４ａは、通常の記憶サブシステム１０４が有する構成要素のほか、リモートコピープログラム１０５ａ、構成情報１０６ａ、リモートコピーキュー１１４の制御情報１０７ａ、リモートコピーの更新情報１０８、リモートコピーされるデータを有する。同様に副記憶サブシステム１０４ｂは、通常の記憶サブシステム１０４の構成要素のほか、リモートコピープログラム１０５ｂ、構成情報１０６ｂ、リモートコピーキュー１１４の制御情報１０７ｂを有する。

【００２６】

構成情報１０６はリモートコピーシステムが有する正記憶サブシステム104a、中間記憶サブシステム104ｃ、副記憶サブシステム104ｂ各々の構成、製品名、アドレス情報に加え、中間記憶サブシステム１０４ｃ内に格納される更新要求（以下ジャーナルとも呼ばれる。）の格納先アドレス及びキュー制御情報の格納先アドレスを有する。構成情報１０６のオリジナル情報は管理コンソール１１７が保持し、中間記憶サブシステム１０４ｃを経由して、または管理ネットワーク１１１を経由して、管理コンソール１１７から各記憶サブシステム１０４へ構成情報１０６の複製が送信され、各記憶サブシステム内に保存される。

【００２７】

正記憶サブシステム104aと副記憶サブシステム104bは構成情報１０６のコピーを持ち、それぞれ記憶サブシステム１０４内のリモートコピープログラム１０５が自記憶サブシステム１０４内に格納されている構成情報１０６を参照してリモートコピーを実行する。

【００２８】

正記憶サブシステム１０４a及び副記憶サブシステム１０４ｂは各々、構成情報１０６から、中間記憶サブシステム１０４c上のリモートコピーキュー領域１１４に格納された更新要求の位置情報、制御情報領域113に格納されたリモートコピーキュー領域１１４に対する制御情報の位置情報を取得する。そして、正記憶サブシステム１０４a及び副記憶サブシステム１０４ｂは各々、取得した位置情報を用いて中間記憶サブシステム１０４c内のリモートコピーキューや制御情報領域にリード若しくはライト要求を送信する。これによって、正記憶サブシステムや副記憶サブシステムは、制御情報１０７を更新したり参照したりし、リモートコピーの更新要求やデータやメッセージを中間記憶サブシステムに書き込んだり、中間記憶サブシステムから読み出したりし、これらの情報を正記憶サブシステムと副記憶サブシステム間で送受信する。

【００２９】

尚、データの整合性を保つために、中間記憶サブシステム１０４c内の各領域（例えば制御情報領域１１３やリモートコピーキュー領域１１４内の各記憶領域）には、所有者となるべき記憶サブシステム１０４を決めておき、データの更新は唯一に決められている所有者である記憶サブシステム１０４から実施する。

【００３０】

中間記憶サブシステム１０４ｃが記憶領域の排他制御機能を有している場合は、当該機能を使用して制御情報格納領域１１３やリモートコピーキュー領域１１４等に格納されるデータの整合性を保つことが望ましい。代表的な排他機能としてはＳＣＳＩのＲｅｓｅｒｖｅコマンドが挙げられる。Ｒｅｓｅｒｖｅコマンドを用いれば、ロジカルユニット（ＬＵ）単位の排他制御が可能であるため、中間記憶サブシステム１０４cは各記憶サブシステム１０４に別々のＬＵを割り当て、排他制御処理を実施すれば良い。この際中間記憶サブシステム１０４cは、各領域の必要サイズに応じてＬＵサイズを設定することが望ましい。尚、本発明においては、排他制御の対象となる記憶領域の単位、排他制御の方式については特に制限はなく、例えばＬＵより小さな単位（たとえばブロック単位等）で排他制御が実行されてもかまわない。

【００３１】

リモートコピーキューの制御情報１０７にはリモートコピーキュー１１４を制御するための情報が含まれる。正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂは各々、自装置が有するリモートコピーキューの制御情報１０７を参照してリモートコピーキュー１１４に対し変更を行い、中間記憶サブシステム１０４cが有する制御情報１１３を更新することによって、変更後の状態をリモートコピーペアの相手方の記憶サブシステム１０４（正記憶サブシステムから副記憶サブシステムへ若しくは副記憶サブシステムから正記憶サブシステム）へ伝える。

【００３２】

正記憶サブシステム１０４ａが有する更新情報１０８とは、正ホスト１０２ａから正記憶サブシステム１０４aが受信するライト要求についての、リモートコピーに関する情報で、各ライトデータの書き込み時刻、書き込み位置、サイズが含まれる。更新情報１０８は正記憶サブシステム１０４aが正ホスト１０２ａからライト要求を受け付けたときに正記憶サブシステム１０４aによって生成され、正記憶サブシステムが中間記憶サブシステム１０４cに送信する更新要求にも含まれるデータである。尚、更新情報１０８は正記憶サブシステム１０４ａが発行したリモートコピーの更新要求が副記憶サブシステム１０４ｂによって取得されたことが確認された後に削除される。中間記憶サブシステム１０４ｃに障害が発生したときには更新情報１０８を使用し，正記憶サブシステム１０４ａと副記憶サブシステム１０４ｃの同期を行う。

【００３３】

尚、ここまでで説明したプログラム１０５は、コンパクトディスクや光磁気ディスクといった可搬記憶媒体を用いて、あるいは、管理ネットワーク１１１を介して、他の装置から各記憶サブシステムが有する記憶媒体にインストールされ、各記憶サブシステムが有するCPUによって実行される。

【００３４】

図２に記憶サブシステム１０４の構成例を示す。

【００３５】

各記憶サブシステム１０４は、コントローラ２０１、一つ以上のディスク装置

２１０を有する。コントローラ２０１は、ホストと接続するためのホストアダプタ２０３、メモリ２０５、ディスク装置と接続するためのディスクアダプタ２０７、プロセッサ２０２、管理系ネットワークと接続するためのネットワークコントローラ２０４を有する。本発明では、記憶サブシステムが有する各構成要素の数には特に制限はないが性能や信頼性の観点から各構成要素は多重化するほうが好ましい。

【００３６】

ホストアダプタ２０３はファイバチャネル等のインタフェース１１０に関するプロトコル制御を行う。ディスクアダプタ２０７はファイバチャネル等のディスクディスクインタフェース２０９に関するプロトコル処理を行う。

【００３７】

メモリ２０５上にはリモートコピー処理に用いられるデータやプログラムが格納されている。即ちメモリ２０５には、リモートコピーを実現するために、リモートコピーのプログラム１０５、リモートコピーの構成情報１０６、制御情報１０７、更新情報１０８、データ２１１が格納される。これらの情報以外にメモリ２０５上には記憶サブシステム１０４を制御するために必要なプログラムや制御情報、ホスト１０２から受信するデータ２１１も格納される。高信頼化のためにメモリ２０５の二重化やメモリ２０５への電源の二重化をしておくことが好ましい。

【００３８】

ネットワークコントローラ２０４は管理系ネットワーク１１１のプロトコル制御を行い、管理コンソール１１７と記憶サブシステム１０４の間の通信を実施する。

【００３９】

ディスクドライブ２１０はディスクインタフェース２０９を介してコントローラ２０１からのリードやライト等のコマンドを受付け、コマンドに従ってリード若しくはライト処理を行う。高信頼化のため，ディスクインタフェース２０９は二重化しておくことが好ましい。一般に記憶サブシステムにおいては、ディスクドライブ２１０を複数組み合わせて冗長構成をとり、複数のディスクドライブの

中に論理的なデバイス（以下、ボリュームともいう。）を作成する。以下、論理的なデバイスであるボリュームを用いて実施形態を説明する。

【００４０】

プロセッサ２０２は記憶サブシステム１０４に関する処理を行う。プロセッサ２０２は内部バス２０８を介しコントローラ２０１内のホストアダプタ２０３、ディスクアダプタ２０７、ネットワークコントローラ２０４と接続しこれらを制御する。更にプロセッサ２０２はメモリ２０５とも内部バス２０８を介し接続し，メモリ２０５内の構成情報１０６や制御情報１０７を用いながらプログラム１０５を実行したり、制御情報１０７を更新したりする。

【００４１】

本実施例においては簡素な内部構成を有する記憶サブシステムの例を説明したが、上述の説明と同等の機能が実現可能であれば、記憶サブシステム104の内部構成に関しては特に制約はない。例えば特開平１０－３３３８３６に開示されている、内部バス２０８の代わりにスイッチを使用した記憶サブシステムを用いても良い。

【００４２】

図３に中間記憶サブシステム１０４ｃのディスクドライブ内に配置される記憶領域の構成例を示す。中間記憶サブシステム１０４ｃのディスクドライブ内には、構成情報１０６が格納される構成情報領域１１２、制御情報１０７が格納される制御情報領域１１３、更新要求が格納されるリモートコピーキュー領域１１４の３つの領域がある。

【００４３】

構成情報１０６には、リモートコピーシステムが有する正記憶サブシステム、中間記憶サブシステム、及び副記憶サブシステム各々の構成、製品名、アドレス情報に加え、リモートコピーキュー領域１１４内の更新要求の格納位置や、制御情報領域１１３内のキュー制御情報１０７の格納先アドレスが含まれる。

【００４４】

制御情報１０７には、リモートコピーキュー１１４を制御・管理するための情報が含まれる。正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂに

宛てた（正→副）制御情報１０７を格納するための正制御情報領域３０１と、副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａに宛てた（副→正）制御情報１０７を格納するための副制御情報領域３０２の二種類の領域をもち、正制御情報領域３０１は正記憶サブシステム１０４ａが、副制御情報領域３０２は副記憶サブシステム１０４ｂが所有権を持つことによって、データの不整合発生を防ぐ。尚、各記憶サブシステムが有する構成情報１０６には、正制御情報領域３０１、副制御情報領域３０２等の各領域のアドレスやサイズ以外に、各領域の所有者を示す情報が含まれており、リモートコピープログラム１０５が構成情報を参照して所有者を確認することで、データの不整合が発生しないように制御を行う。ＳＣＳＩのＲｅｓｅｒｖｅコマンドのように物理的な排他が可能な場合はより堅牢なシステムが構成可能になる。この場合には、中間記憶サブシステム１０４cが論理ユニット（ＬＵ）ごとに正制御情報領域、副制御情報領域を割り当て、ＬＵごとに排他制御を行うことになる。

【００４５】

リモートコピーキュー１１４に格納される情報には、大別してメッセージとデータの２種類の情報があるため、各情報を格納するためにリモートコピーキュー１４４にはメッセージ領域とデータ領域の2種類の領域がある。以下、2種類の領域を特に区別しない場合はキュー１１４として説明する。キュー１１４には、制御情報１０７と同様、正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂに宛てた（正→副）情報と副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａへ宛てた情報の両方が格納される。従って、キュー１１４には少なくとも、正記憶サブシステム104aから副記憶サブシステム104bに宛てて発行される、メッセージを格納する正メッセージ領域３０３とデータを格納する正データ領域３０５、副記憶サブシステム104bから正記憶サブシステム104aに宛てて発行される、メッセージを格納する副メッセージ領域３０４とデータを格納する副データ領域３０６の４つの領域が存在する。尚、リモートコピーキュー１１４に存在する記憶領域の数はリモートコピーシステムの構成に依存し、例えば副記憶サブシステムを複数有するリモートコピーシステムであれば、リモートコピーキュー１１４に存在する記憶領域の数は４より多くなることもある。

【００４６】

制御情報領域１１３と同様キュー領域１１４においても、正記憶サブシステム１０４aから副記憶サブシステム１０４bに宛てた情報（メッセージ若しくはデータ）が格納される領域は正記憶サブシステムが所有者となっており、副記憶サブシステム104bから正記憶サブシステム104aに宛てた情報は副サブシステムが所有者となっている。

【００４７】

図４に中間記憶サブシステム１０４ｃのディスクドライブに格納されるデータの構造例を示す。

【００４８】

中間記憶サブシステム１０４ｃのディスクドライブ２１０は、通常のディスクドライブであり、中間記憶サブシステム１０４cも通常の記憶サブシステムであるため、ディスクドライブ２１０に書き込まれた情報は他の装置から上書きすることができる。従ってディスクドライブ２１０内の記憶領域に書き込まれたデータが、所有者の論理的障害(プログラムのバグ)または、当該記憶領域の所有者以外の他の装置によって上書きされ、データがエラーになる（不整合となる）可能性がある。中間記憶サブシステム１０４ｃでこれらのエラーを検知可能であればエラーに対する対応が可能であるが、所有者以外のホスト１０２又は記憶サブシステム１０４からのデータ上書き、論理エラー等を中間記憶サブシステムが検知するのは困難な場合もある。

【００４９】

そこで本実施形態においては、中間記憶サブシステム１０４ｃに書き込まれるデータに保護データを付加することにより、正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂの間で中間記憶サブシステム１０４cを介して授受されるデータの保障を行う。

【００５０】

保護データが付加されたデータの例を図４に示す。尚本実施形態においては、中間記憶サブシステム１０４ｃはＳＣＳＩ等のブロックデバイスを有する記憶サブシステムであると仮定している。通常ブロックは５１２バイトのサイズを有し

、中間記憶サブシステム１０４cが有するディスクデバイスは、ブロック単位でアクセスされる。

【００５１】

図4（ａ）は、中間記憶サブシステム１０４ｃが有するディスクデバイスの各ブロック４０１の最後の領域を保護データ領域４０２とした例である。保護データ４０２を各ブロックの最後に書き込むため、従来はブロックの最後の領域に格納することのできた元データが次のブロックの先頭位置（４０３）に格納されることになる。従来次のブロックの先頭であった位置（４０４）は記憶領域４０３の後になるため、この方法を用いるとデータの格納位置が順次後ろにずれていく。

【００５２】

このため従来一ブロック内に格納されていたデータが、保護データ領域を設けたことにより複数ブロックにまたがって格納されることになる。すると、一ブロック分のデータを中間記憶サブシステム１０４に格納するために複数ブロックを更新する必要が生ずる。ブロックデバイスはブロック単位でアクセスされるため、この場合は関連するブロックをすべてディスクデバイスからコントローラ内のメモリ２０５上に読み込み、データを更新し、これをディスクデバイス２１０に書き込む必要がある。しかし、本発明において中間記憶サブシステム１０４ｃに格納されるデータは、一度書き込まれるのみで、同一データの読み書きが行われないため図4(a)に示す方法を採用しても問題は少ない。

【００５３】

図４（ｂ）は中間記憶サブシステム１０４ｃのブロックサイズを保護データ分だけ増加させ、増やした領域に保護データ４０２を格納する例である。この方法を用いるとブロックアドレスを変更することなく、保護データ４０２が追加されたデータをディスクデバイスに格納することが可能である。

【００５４】

図５に保護データ４０２の一例を示す。図４において説明したとおり保護データ４０２は中間記憶サブシステム１０４ｃに格納される各ブロックに付加されている。保護データ４０２には大別して２種類のデータが含まれる。一つは要求の

識別子５０１と要求内の通し番号（シーケンス番号）５０２とを有する論理的な情報である。他の一つはブロックデータのエラーをチェックするための誤り検出符号５０３である。前者は論理的な誤りや他者によるデータ改ざんを検出するために用いられ、後者はブロック自体のデータ障害を検出するために用いられる。誤り検出符号５０３にはパリティ、チェックサム、ＥＣＣ、ＣＲＣ、ハミング符号等あるが本発明においては誤り検出及び実装が可能であればどの方式の誤り検出符号を用いてもかまわない。

【００５５】

保護データ４０２はデータを生成する記憶サブシステム１０４のホストアダプタ２０３にてデータに付加され、受信側の記憶サブシステム１０４のホストアダプタ２０３で保護データを用いたデータのチェックが行われる。

【００５６】

図６に中間記憶サブシステム１０４ｃに格納されるデータの論理構造の一例を示す。図６は、あるリモートコピーグループ内の正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ送信される更新要求に関する情報を示している。

【００５７】

構成情報領域112内に格納されている構成情報１０６には静的な構成の情報が含まれ、構成情報１０６はリモートコピーグループの構成変更時に更新される。各記憶サブシステム１０４は構成情報１０６を中間記憶サブシステム１０４ｃの構成情報格納領域112から取得して自装置内に展開し、構成情報からリモートコピーに関する情報およびリモートコピーキュー領域１１４に関する構成情報を取得する。そして、これらの情報を利用して各記憶装置サブシステムはリモートコピーおよびキュー制御を実行する。

【００５８】

本実施例では中間記憶サブシステム１０４ｃを介して各記憶サブシステム１０４へ構成情報１０６を配布しているが、構成情報１０６は管理コンソール１１７から各記憶サブシステム１０４へ直接送付しても良い。

【００５９】

図６では一つのリモートコピーグループに関する構成情報１０６を示している。

【００６０】

通常リモートコピーグループとは一つ以上のボリュームから構成され、各ボリューム間の整合性（コンシステンシ）を保つグループとして定義される。リモートコピーグループに対するサスペンド、再同期等の操作は、ボリュームペア単位の他、グループ単位でも実行可能である。リモートコピーグループは一つの正記憶サブシステム１０４ａと一つ以上の副記憶サブシステム１０４ｂで構成され、識別子（グループＩＤ）６０１が付けられている。従って構成情報１０６には、グループID６０１、リモートコピーグループに属する正記憶サブシステムのID６０２と副記憶サブシステムのID６０４、副記憶サブブシステム１０４ｂの個数６０３が含まれる。さらに構成情報１０６には、キュー領域１１４内に存在する正メッセージ領域や正データ領域等の領域（以下、キューとも呼ぶ）の数６０５と各キューの個別情報６０６が格納される。

【００６１】

図３に示したように、正記憶サブシステム１０４と副記憶サブシステム１０４が一対一の基本構成では、メッセージとデータに関して双方向のキューが生成されるため、一つのリモートコピーグループに関して４つのキューが存在する。図６では正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ宛てて送信される正データのキューに関する個別情報を示している。逆方向のキューに関しても、正・副が入れ替わるだけで、同様の制御が実施される。

【００６２】

キューの個別情報の中にはキューの識別子６０７、メッセージ又はデータのいずれに関するキューかを識別するためのキュー種別６０８、中間記憶サブシステム１０４ｃのキュー領域１１４内における当該キューの先頭位置６０９、キューのサイズ６１０、正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ送付される制御情報１０７に関する正情報６１１、副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａへ送付される制御情報１０７に関する副情報６１２が含まれる。

【００６３】

尚、構成情報領域１１２とキュー領域１１４とが別々の記憶サブシステム１０４に存在する場合には、先頭位置６０９として、キュー領域が配置される記憶サブシステムの識別子、当該キューのＬＵ等の論理アドレス等が含まれる。副記憶サブシステムが複数個存在する場合にはこれらの情報も複数個必要になる。

【００６４】

制御情報１０７に関する情報６１１や６１２には、制御情報の発行元記憶サブシステム１０４の識別子６１３、制御情報領域１１３内での制御情報の格納位置を示す先頭位置６１４、制御情報のサイズ６１５が含まれる。各記憶サブシステム１０４は構成情報１０６内の制御情報に関する情報（例えば６１１や６１２）を元に、制御情報１０７が格納されている記憶領域を特定して制御情報１０７をやり取りする。

【００６５】

一つのキューに対応する制御情報１０７としては、一つの正制御情報６３１と一つ以上の副制御情報６３２がある。各制御情報１０７は予め定められた唯一の記憶サブシステム１０４によって更新されるが、制御情報の参照は複数の記憶サブシステム１０４から可能である。

【００６６】

各制御情報１０７は、リモートコピーのグループ識別子６１６、キューの識別子６１７、当該制御情報１０７の所有装置を示す所有装置識別子６１８、リモートコピー処理の進捗状況を示すためリモートコピー中データの先頭位置を示す先頭位置６１９、リモートコピー中のデータのサイズ６２０、及びハートビート６２１を有する。

【００６７】

グループＩＤ６１６、キューＩＤ６１７、所有装置識別子６１８は、制御領域が論理的に正しいかをチェックするために使用される。先頭位置６１９及びサイズ６２０には各々、リモートコピー処理中のデータが格納されているキュー領域内の記憶領域の先頭位置と、当該記憶領域のサイズが格納される。更に各記憶サブシステム１０４はハートビート６２１に定期的に情報を書き込むことで、自装

置が稼動中であることを他の装置に通知する。言い換えれば、ハートビート６２１が更新されているか否かを確認することで、当該ハートビートに対応する記憶サブシステムが稼動中であるか否かを他の装置から判別できる。尚、ハートビート領域６２１に書き込まれる情報はカウンタ値やタイマなど時間に伴い変化する値であることが望ましい。

【００６８】

正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂの制御、動作は同期していないため、中間記憶サブシステム１０４ｃの制御情報領域113に格納されている正制御情報631の先頭位置619及びサイズ620によって示されるデータと、副制御情報632の先頭位置619及びサイズ620によって示されるデータと、正記憶サブシステム１０４ａが有する制御情報107aの先頭位置626及びサイズ627によって示されるデータと、副記憶サブシステム１０４ｂが有する制御情報107bの先頭位置及びサイズによって示されるデータとは同一とは限らない。

【００６９】

図６を用いて正制御情報631、副制御情報632、正記憶サブシステム104aが有する制御情報107a、及び副記サブシステム104bが有する制御情報107bの関連を説明する。

【００７０】

キュー領域１１４にはＤａｔａ１～Ｄａｔａ６が格納されているものとする。また、図６において中間記憶サブシステム１０４ｃ上の正制御情報６３１は、先頭位置６１９がＤａｔａ1を示し、サイズ620がData1～Data4までのサイズを示しており、この結果Data1を先頭にＤａｔａ４までを処理中の要求（処理中のデータ）として示しているものとする。副制御情報６３２は、先頭位置６１９がＤａｔａ３を示し、サイズ620がData3分のサイズを示しており、この結果Data3を処理中の要求（処理中のデータ）として示しているものとする。

【００７１】

更に、正記憶サブシステム１０４ａが有する制御情報１０７ａは先頭位置６２６がＤａｔａ１を示し、サイズがData1からData6までのデータのサイズを示しているものとする。即ち結局制御情報１０７aは、Data1からData6までを示してい

ることとなり、正記憶サブシステム１０４ａはＤａｔａ１～Ｄａｔａ６をリモートコピー中のデータとして認識しているものとする。正記憶サブシステム104aが次のデータをキュー領域114に書き込む場合には、制御情報107aの次位置628が示す、Ｄａｔａ６の次の位置からデータを書き込む。

【００７２】

一方副記憶サブシステム１０４ｂはＤａｔａ１、及びData２をすでに中間記憶サブシステム104cから取得しており、副記憶サブシステム104bが有する制御情報１０７bの先頭位置６２９はＤａｔａ３を、サイズ６３０はData3分のデータ量を示しているものとする。従って、副記憶サブシステム104bは、Data3を処理中の要求（処理中のデータ）と認識している。

【００７３】

尚、副記憶サブシステム１０４bは、中間記憶サブシステム１０４ｃ上の正制御情報６３１を参照することによって、正記憶サブシステムがＤａｔａ１を先頭６２４にＤａｔａ４までを処理中の要求（処理中のデータ）として認識している、と判断する。従って、副記憶サブシステム104bは、既に処理したData1、及びData2と、処理中のData3のみならず、Data4も処理すべきことを、中間記憶サブシステム104cに格納されている正制御情報631を参照することによって、認識することができる。

【００７４】

Ｄａｔａ５及びＤａｔａ６は、中間記憶サブシステム１０４ｃ上の正制御情報631が更新された後に正記憶サブシステム１０４ａからキュー領域１１４に追加された情報である。キュー領域へのデータの書き込みと、正制御情報631の更新とは非同期で実行されるので、図6においては、Data5及びData6のキュー領域114への追加に関しては正制御情報631にはまだ更新されていない。最新の制御情報は正記憶サブシステム１０４ａ内に制御情報107aとして存在し、一定時間後、正記憶サブシステム１０４ａから中間記憶サブシステム１０４ｃへの正制御情報６３１更新契機が訪れた際に、正制御情報６３１が最新情報に更新される。

【００７５】

次に、正記憶サブシステム104aが中間記憶サブシステム１０４ｃ上の正制御情

報631やキュー領域内のデータ（以下、ジャーナル若しくは更新要求ともいう）、又は正記憶サブシステムが有する制御情報107aを更新する処理について説明する。

【００７６】

正制御情報631及びキュー領域内のデータは、正記憶サブシステム１０４ａが、ホスト１０２からのライト要求と同期または非同期にて更新する。

【００７７】

正記憶サブシステム１０４ａが有する制御情報107aの次位置628は、正記憶サブシステム104aがキュー領域１１４へ更新要求を送付するのと同期して、正規億サブシステム１０４ａ内で更新される。正記憶サブシステム104aが有する制御情報107aの先頭位置６２６は、中間記憶サブシステム１０４cが有する副制御情報632の先頭位置619に基づいて、正記憶サブシステム104aによって、ホストからのライト要求とは非同期にて更新され，トラフィックを削減するために複数ライト分を更新可能である。正記憶サブシステム104aの制御情報107aのサイズ６２７に関しては、先頭位置６２６と次位置６２８各々の更新に合わせて更新される。

【００７８】

正記憶サブシステム１０４ａは定期的に中間記憶サブシステム１０４ｃの制御情報領域１１３に格納されたデータにアクセスする。このとき正記憶サブシステム104aは、副制御情報６３２を取得し（リードし）、副制御情報６３２の先頭位置６１９を参照して、正記憶サブシステムが有する制御情報１０７aの先頭位置６２６を変更する。例えば上述の例においては、制御情報107aの先頭位置626は現在Ｄａｔａ１を示しているが、副制御情報632の先頭位置619はＤａｔａ３を示しているため、正記憶サブシステムは中間記憶サブシステム104cの制御情報領域113にアクセスした際に、副制御情報632を取得して、自装置の制御情報107aの先頭位置６２６をＤａｔａ３に更新する。

【００７９】

この際正記憶サブシステム１０４ａに格納されている更新情報１０８もあわせて削除可能となる。即ち、正記憶サブシステムは副制御情報632を参照することによって、Data１及びData2については副記憶サブシステムによって取得済みで

あることが認識できるので、Data１及びData２についての更新情報１０８は削除できる。削除のタイミングは、先頭位置626の更新時点以降であればどのタイミングでもよい。

【００８０】

また、正記憶サブシステム104aは、制御情報107aの正サイズ627も、正記憶サブシステム104aの制御情報107aが有する先頭位置626と次位置628から算出して更新する。係る処理によって制御情報107aの正サイズは、Data3からData6までのデータ量を示すことになる。

【００８１】

以上の処理により、正記憶サブシステムが有する制御情報107aは最新のデータに更新されるので、正記憶サブシステムはこれを中間記憶サブシステムの制御情報領域113に書き込むことにより、正制御情報631を更新する。

【００８２】

次に副記憶サブシステム１０４ｂが中間記憶サブシステム１０４ｃの制御情報領域113にアクセスし、副制御情報６３２を更新する処理について説明する。

【００８３】

副記憶サブシステム１０４ｂは一定時間間隔で中間記憶サブシステム１０４ｃから正制御情報６３１を取得し、正制御情報631が示す更新要求のうち、未だ取得していない更新要求をキュー領域１１４から取得する。その後、副記憶サブシステム104bは、自装置が有する制御情報107ｂと副制御情報６３２を更新する。

【００８４】

上述の例においては、最初正制御情報６３１はＤａｔａ１からＤａｔａ４までを処理中のデータとして示し、副制御情報６３２の先頭位置619はＤａｔａ３を示している。従って、副記憶サブシステム104bは正制御情報631を参照することによって、Data３とData4を自装置が未だ取得していないことを認識する。そこで副制御サブシステムはＤａｔａ３とＤａｔａ４をキュー領域114から取得した後、自装置１０４ｂが有する制御情報107b及び中間記憶サブシステム１０４ｃが有する副制御情報632を更新する。即ち、制御情報107bの先頭位置629と副制御情報632の先頭位置619は共にData4を、制御情報107bのサイズ630と副制御情報632

のサイズ620は共に0を示すよう、制御情報が更新される。

【００８５】

尚、副記憶サブシステム１０４bの稼働率が高い等、何らかの理由で副記憶サブシステム１０４ｂがキュー領域114から更新要求を取得しなかった場合には、副記憶サブシステム１０４bは制御情報の更新のみを行う。この場合、制御情報107bの先頭位置629と副制御情報632の先頭位置619は共にData3を、制御情報107bのサイズ630と副制御情報632のサイズ620は共にData3とData4を合計したデータ量を示すよう、制御情報が更新される。

【００８６】

図７に、正記憶サブシステム若しくは副記憶サブシステムから発行され、キュー内に格納される各要求７２１のデータ構造例を示す。

【００８７】

要求の先頭にはヘッダ情報７２２が格納され、末尾には末尾情報７２４が格納される。ヘッダ情報７２１には要求の属性７０１、グループＩＤ７０２、キューＩＤ７０３等の要求の属性と、要求の個別情報が格納される。要求の属性は論理エラーチェックのために用いられる。要求の個別情報には時系列情報と位置情報の二種類の情報がある。時系列情報としてはキュー内における要求の通番（要求ＩＤ）７０４、正記憶サブシステム１０４ａまたはホスト１０２にて付加されたタイムスタンプ７０５があり、副記憶サブシステム１０４ｂ内で要求を時系列順に整列し、要求の抜けをチェックするときに用いられる。位置情報にはボリュームのＩＤ７０６、ボリューム内アドレス７０７、要求のサイズ７０８があり、これらの位置情報を元に要求に含まれるデータが副記憶サブシステム１０４ｂ内に格納される。尚、要求に含まれる位置情報は、ホストから受信するライト要求に含まれる位置情報と同じものである。サイズ７０８はホストから受信するライトデータのサイズを示し、このサイズ７０８にヘッダ情報７２１と末尾情報７２４の各固定サイズを足したサイズが要求のサイズとなる。

【００８８】

データ７２３はホストから受信するライトデータである。

【００８９】

末尾情報７２４にはヘッダ情報７２２に加え誤り検出符号７０９が含まれる。誤り検出符号７０９はヘッダ情報７２２とデータ６３３とを通して計算され、要求全体に関する誤りの検出を行う。要求の誤り検出符号７０９は図５で示した保護データ４０２と併用することでより高信頼なリモートコピー処理を実現する。

【００９０】

このように、ヘッダ情報７２２や末尾情報７２４等の制御情報とライトデータ７２３を連続した領域に書き込むことで、ホストからライト要求を受信した際にこれに対応して中間記憶サブシステム１０４ｃへ発行される要求は、連続領域に対するライト要求となり、中間記憶サブシステム１０４ｃに対する書き込みは一度行うだけでよい。

【００９１】

尚、上述のように各要求にはホスト１０２から受信する位置情報等の制御情報とライトデータとが双方含まれており、ジャーナルとも呼ばれる。

【００９２】

図８は本実施形態における初期設定の手順の一例を示すフローチャートである。最初に管理者は管理コンソール１１７を介して、リモートコピーシステムを構成する装置の情報、各記憶サブシステム内のボリューム等の記憶サブシステム１０４の情報、リモートコピーシステム内に存在するホスト１０２上で稼動するアプリケーションの情報や、ホスト１０２が使用しているボリュームの情報を取得する（８０１）。

【００９３】

次に管理者は、管理コンソールが収集したこれらの情報を元に、リモートコピーのボリュームペアや、ボリュームペアを集めたコンシステンシグループ等を決定し、各記憶サブシステム１０４のリモートコピーの構成情報１０６を作成して管理コンソール１１７に入力する（８０２）。構成情報１０６については、管理コンソール１１７がオリジナル情報を保持し、各記憶サブシステム１０４へは中間記憶サブシステム１０４ｃを経由して、または管理ネットワーク１１１を経由してコピーが送信される（８０３）。

【００９４】

各記憶サブシステム１０４に構成情報１０６が設定され、リモートコピーのペアが確立されたり、コンシステンシグループが生成された後、正記憶サブシステム１０４ａから中間記憶サブシステム104cを介して副記憶サブシステム１０４ｂへのデータの初期コピーが実行され、リモートコピー処理が開始される（８０４）。

【００９５】

図９に、本実施形態において正記憶サブシステム１０４ａがホスト102からライト要求を受信した場合に実行する更新処理の一例を示す。

【００９６】

正記憶サブシステム１０４ａはホスト１０２からライト要求を受け付ける（９０１）。そして正記憶サブシステム104aは、ライト要求から関連するボリューム、アドレスを算出し、以下に示すチェック処理を行う。まず正記憶サブシステムは、ライト要求の対象であるボリュームに関して、リモートコピー属性が指定されているかどうかを調べ、リモートコピー属性が指定されていない場合は、更新データ（以下ライトデータともいう。）をホスト１０２から受領し（９１１）、ホスト１０２へライト要求に対する処理の終了を報告して（９０９）処理を終える。

【００９７】

リモートコピー属性が指定されている場合は、ステップ９０３へ進む。正記憶サブシステム104aは、ライト対象ボリュームに対応するリモートコピーペアの状態をチェックし、正常にペアが構成されている状態（ペア状態）か、ペアが切り離されている状態か（サスペンド）をチェックする（９０３）。

【００９８】

サスペンド状態になっている場合には差分情報を記録し（９１０）、更新データをホスト１０２から受領して（９１１）、ホスト１０２へ更新処理の終了を報告し（９０９）、処理を終える。尚差分情報は、任意のサイズの位置情報を一ビットに対応させたビットマップの形式で保持される。ビットマップで差分情報を保存する場合には，同一アドレスへの上書きは同じビットで示されるため、差分情報を格納するために確保しなければならない領域が少なくてすむ反面、ホスト

から受信するライト要求の順序性（受信順序）を保存することはできない。順序性を保持したい場合には更新ログ等、各ライト要求の受信時刻、ライト対象記憶領域のアドレス、ライトデータのサイズ、ライトデータを保持する必要があり、ビットマップで差分情報を管理する場合より多くの記憶容量が必要になる。通常サスペンドは長時間にわたることが多いためビットマップで差分情報を保持することが多い。ビットマップは記憶サブシステム104のメモリ２０５上に格納してもよいが、ディスク２１０上に格納してもよい。

【００９９】

ペア状態である場合には、正記憶サブシステム104aは中間記憶サブシステム１０４ｃ上の記憶領域の有無をチェックする。具体的には正記憶サブシステム１０４ａが保持する構成情報１０６ａと制御情報１０７ａから中間記憶サブシステム104c内に存在するキュー領域114の残りの記憶容量を計算する（９０４）。

【０１００】

通常中間記憶サブシステム１０４ｃには、ホスト１０２の負荷で、即ちホストからのライト要求及びライトデータによってキュー領域114が全て埋ってしまうことのないよう、十分大きなキュー領域114を用意しておく。ただし予期しない負荷が発生する場合もあり、この場合は中間記憶サブシステム１０４ｃのキュー領域を使い切ってしまう。

【０１０１】

図９では中間記憶サブシステム104cのキュー領域に空き容量がない場合には、正記憶サブシステム104aがホスト１０２へビジー報告を返し、ライト要求を受け付けない（９１２）。ホスト１０２へビジーを返送する代わりに、正記憶サブシステム104aは処理（９１２）において中間記憶サブシステムに障害が発生していると判定し、リモートコピーの状態をサスペンド状態としてもよい。この場合には正記憶サブシステム104aは、図１１に示す障害処理を実施し、リモートコピーの状態をサスペンドへ移行し、ホストから更新データ（ライトデータ）を受信してビットマップを更新し、ホストに対し通常の終了報告を行う。

【０１０２】

ステップ９０４にて中間記憶サブシステム104cに空き容量があると判断された

場合、即ちステップ９０２から９０４までの処理にてリモートコピー処理が可能であると判定された場合には、正記憶サブシステム104aはステップ９０５以下の処理を実施する。まず正記憶サブシステム104aはホスト１０２からライト要求に対応するライトデータを受領する（９０５）。

【０１０３】

次に正記憶サブシステム104aは、該ライト要求に対応するリモートコピーの更新要求を作成し、ライトデータと合わせて図７に示すような更新要求を作成してこの更新要求を中間記憶サブシステム１０４ｃのキュー領域114の次位置以降に書き込む（９０７）。このとき更新要求を書き込もうとするとキュー領域１１４の最末尾を超えてしまう場合は、超える分の続きのデータはキュー領域１１４の先頭から書き込む。尚、更新要求をキュー領域114に書き込む際には、副記憶サブシステム１０４ｂが有する制御情報107bが示す先頭位置629を越えないように正記憶サブシステム104aは制御する必要がある。

【０１０４】

次に正記憶サブシステム104aは、キュー領域114に書き込んだ更新要求のサイズ分だけ、制御情報107aの次位置628、サイズ627を変更し（９０８）、ホスト１０２へ終了を報告した後（909）処理を終える。

【０１０５】

図１０に、正記憶サブシステム１０４ａと中間記憶サブシステム１０４ｃとの間で実行される制御情報の更新処理の一例を示す。まず正記憶サブシステム１０４ａは中間記憶サブシステム１０４ｃから、副記憶サブシステム１０４ｂの制御情報である副制御情報６３２を取得する（１００１）。

【０１０６】

そして正記憶サブシステムは、副制御情報６３２内のハートビートを参照し副記憶サブシステム１０４ｂが稼動しているか否かをチェックする。具体的には、正記憶サブシステム１０４ａは過去に取得した副記憶サブシステムのハートビート情報の内最も新しいハートビート情報を正記憶サブシステム１０４ａ内のメモリ２０５に保存しておき、中間記憶サブシステム１０４ｃから新たに副制御情報32を取得した際には、取得した副制御情報632中のハートビート情報と、メモリ

に格納されているハートビート情報とを比較して、チェックを行う。2つのハートビート情報が同じ値であれば副記憶サブシステムは稼動していないこととなる。

【０１０７】

尚、制御情報の更新処理は正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂとの間では非同期に実施されるため、一度ハートビート情報が更新されていないと判断されただけで副記憶サブシステム104bが稼動していないと判断するのではなく、数回連続してハートビート情報の更新がない場合や一定時間以上更新がない場合等に幅を広げて、副記憶サブシステム104bの稼動状況を判断する。たとえば、副制御情報６３２の更新周期が１秒である場合は、５秒以上ハートビート情報の更新がない場合に副記憶サブシステムが稼動していないと判断する等とする。

【０１０８】

ステップ１００２において副記憶サブシステムが稼動していないと判断された場合には、正記憶サブシステム104aが図11に示す障害時処理を行う（１００６）。

【０１０９】

ステップ１００２において副記憶サブシステムが稼動していると判断された場合には、正記憶サブシステム104aは中間記憶サブシステム104cの制御情報領域113に格納されている副制御情報632を参照して、自装置が有する制御情報107aを更新する。即ち、制御情報107aの正先頭位置626を副制御情報632の副先頭位置619に合わせ、制御情報107aの正サイズ627は更新後の正先頭位置626から正次位置628までのサイズとする。そして正記憶サブシステム１０４ａの新制御情報１０７ａを用いて、中間記憶サブシステム１０４ｃの制御情報領域113に格納されている正制御情報631を更新する（１００３）。

【０１１０】

次に正記憶サブシステムは更新情報108を廃棄する。副記憶サブシステム１０４ｂで新たに更新された更新要求、即ちステップ１００３における変更前の先頭位置から変更後の先頭位置までに存在する更新要求は、既に副記憶サブシステム

によって取得されている。従ってこの更新要求に対応する更新情報については、正記憶制御装置にて保持する必要がないため、正記憶サブシステムはこの更新情報108を廃棄可能とし、任意の時間に破棄する（１００４）。

【０１１１】

更に正記憶サブシステム104aは、ステップ１００１からステップ１００４の処理を一定間隔で実施するため一定時間（例えば１秒間）待機した後（１００５）、ステップ１００１から再び処理を繰り返す。

【０１１２】

図１１はリモートコピーシステムに障害が発生した際に正記憶サブシステムが実行する処理の一例を示している。

【０１１３】

正記憶サブシステム104aは障害を検出すると（1101）、障害が発生している部位を特定して（1102）、管理コンソール１１７に対して障害を報告する（1103）。障害の報告を受けた管理コンソール117は、管理者からの指示に基づいて障害部位を閉塞させる（１１０４）。尚、障害部位の閉塞は管理コンソール117以外にも障害を検出した正記憶サブシステム１０４からも行う。

【０１１４】

障害部位を閉塞した後、正記憶サブシステムはリモートコピーの通信経路の状態を取得し、交替パスが存在するか否かをチェックする（１１０５）。交替パスが存在する場合にはリモートコピーの処理が継続可能であるため、正記憶サブシステム１０４は交替パスへのリモートコピー用パスの切り替えを行った後に障害処理を終える。

【０１１５】

交替パスが存在しない場合には正記憶サブシステムはリモートコピーペアの状態を変更する。尚ステップ９０４において、正記憶サブシステム１０４aが中間記憶サブシステム１０４ｃのキュー領域に空き領域がないと判断した場合であって、キュー領域に空き領域がない状態を障害の発生とみなす場合にも、交替パスが存在しない場合として扱う。

【０１１６】

交代パスが存在しない場合、正記憶サブシステム１０４が障害を検出していれば、自らペア状態を変更し(1106)、差分情報（ビットマップ）に副記憶サブシステム１０４bには格納されていないライトデータの位置情報を登録する（1107）。尚、別途管理コンソール１１７がペアの状態変更、差分情報（ビットマップ）の作成処理を正記憶サブシステムに指示しても良い。

【０１１７】

ステップ1106によって、ペアの状態は障害検出前の正常状態（ペア状態）からペアが切り離されている状態（サスペンド）へ変更される。また、ビットマップは正記憶サブシステム１０４a上のメモリ２０５またはディスクドライブ２１０上に作成される。尚、ビットマップ作成開始時点の時刻、各グループの更新要求に関する要求番号等も記憶サブシステムに記憶される。

【０１１８】

また、正記憶サブシステム１０４ａは未反映のリモートコピー更新要求を先頭にビットマップを作成する。例えばリモートコピーシステムが図６の状態にあった時点でサスペンド状態となった場合には、正記憶サブシステム１０４ａはビットマップを作成してＤａｔａ１～６に関する差分情報をビットマップに格納し、これを初期状態とした後に、それ以後ホスト１０２から受信したライト要求についてビットマットを作成する（910）。

【０１１９】

障害発生後も正記憶サブシステム１０４ａに対して正ホスト１０２ａからの業務が継続される場合は、ホスト102aから正記憶サブシステム１０４ａがライト要求とライトデータを受信する度に、ステップ1107の処理によって、正記憶サブシステム１０４ａ上のビットマップへ差分情報が格納される。

【０１２０】

尚、上記のステップ１１０1からステップ１１０６に示す処理は処理時間が十分短いためホスト１０２のＩＯ要求を正記憶サブシステムが受け付け、処理を継続したまま処理可能である。またステップ１１０７に示す処理も全リモートコピーペア同時に実施すると時間を要するが、個々のリモートコピーペアに対しては十分短い時間で処理可能である。ステップ1107に示すビットマップの更新処理は

ホスト１０２からのＩＯ要求に同期して実行することによりリモートコピーペア間では分散処理が可能となるので、ホスト１０２のＩＯ処理を継続したまま処理できる。

【０１２１】

図１２は副記憶サブシステム１０４ｂが制御情報を取得し、これを更新する処理の一例を示す図である。

【０１２２】

まず副記憶サブシステム１０４ｂは構成情報１０６に基づいて、正制御情報631を中間記憶サブシステム１０４ｃの制御情報領域113から取得する（１２０１）。次にステップ１００１と同様、制御情報取得が正常に行われたか、ハートビート情報が更新されているか等をチェックする（１２０２）。ステップ１２０２において障害を検出した場合には、副記憶サブシステムは図１３に示す障害時処理を実施する（１２０９）。

【０１２３】

正常な状態の場合、即ち障害が検出されなかった場合には、副記憶サブシステムはまず更新要求の有無を調べる（１２０３）。副記憶サブシステムは、取得した正制御情報631の正先頭位置619と正サイズ620で示されるリモートコピー処理中の要求と、自装置が有する制御情報107bの先頭位置629とサイズ630で示されるリモートコピー処理中の要求との差分から、前回自装置が制御情報107bを更新した時点以降にキュー領域114に追加された更新要求を把握することができる。

【０１２４】

ステップ１２０３において新たな更新要求（即ち、副記憶サブシステムが前回制御情報107bを更新した後、新たにキュー領域に追加された更新要求）がない場合には、副記憶サブシステムは制御情報更新処理を終え一定期間待機する（１２０８）。

【０１２５】

更新要求がある場合には、副記憶サブシステム104bは当該更新要求をキュー領域114から取得する（１２０４）。副記憶サブシステム１０４ｂが取得可能なデータの容量が十分大きい場合には、副記憶サブシステム104bは取得すべき更新要

求を全て一度にキュー１１４より取得する。

【０１２６】

その後、副記憶サブシステム１０４ｂ内にて更新要求の内容を解析し、更新要求に従って正記憶サブシステムから送信されたデータを副記憶サブシステム１０４ｂ内のボリュームへ反映する（１２０５）。即ち、更新要求に含まれるアドレス７０７が示す記憶領域に、更新要求に含まれるデータ７２３を格納する。

【０１２７】

その後更新を行ったデータに対応する更新要求の分だけ、副記憶サブシステムが有する制御情報１０７ｂを更新し（具体的には先頭位置629と副サイズ630を更新し、）（１２０６）、更新後の制御情報107bの内容を用いて中間記憶サブシステム１０４ｃの制御情報領域１１３に格納されている副制御情報632を更新する（１２０７）。

【０１２８】

尚、ステップ１２０３において更新要求のサイズが大きく、更新要求を一度に取得することができない場合、副記憶サブシステム104bは更新要求を分割して取得し、処理を行う。まず副記憶サブシステム１０４ｂは更新要求取得のための領域を用意し更新要求を先頭位置から取得する。この領域は副記憶サブシステム１０４ｂが用意可能な領域、例えば１０ＭＢ等となる。

【０１２９】

尚、この記憶領域に格納される最後の更新要求に関しては、ほとんどの場合取得できるのは最後の更新要求の途中までである。例えば１０ＭＢの領域が副記憶サブシステムに確保されており、ここに既に９ＭＢ分の更新要求が取得済みとなっている場合を考える。副記憶サブシステムは次の更新要求を９MBの位置から自装置内の領域に読み込むわけだが、次の更新要求のサイズが２ＭＢであった場合は、当該更新要求のうち最初の１ＭＢ部分は取得できるが、残りの部分は取得できない。

【０１３０】

このような場合副記憶サブシステムは、最初の１ＭＢ部分のみ中間記憶サブシステムから更新要求を取得し、当該部分に対応するヘッダ情報を解析して先頭部

分に格納されているデータのみ副記憶サブシステム内のボリュームへ反映する。そして、処理した更新要求の先頭部分に関してのみ、制御情報１０７bの副先頭位置６２９、副サイズ６３０を変更して、中間記憶サブシステム１０４ｃ上の制御情報１１３を更新する。尚、更新要求の残りの部分に関しては、副記憶サブシステムは次の更新要求取得時に処理する。

【０１３１】

図１３は障害発生時に副記憶サブシステム１０４ｂにおいて実行される処理の一例を示す図である。障害時には副記憶サブシステム１０４ｂは可能な限り中間記憶サブシステム１０４ｃ上に残留する更新要求を取得する。正記憶サブシステム１０４ａと中間記憶サブシステム１０４ｃは同期しているため、正記憶サブシステム１０４ａのみに障害が起きた場合であって、中間記憶サブシステム１０４ｃにはアクセス可能な場合には、副記憶サブシステム１０４ｂは中間記憶サブシステム１０４cに格納されている更新要求を取得することにより、データ消失を防ぐことができる。

【０１３２】

副記憶サブシステム104bはまず、正記憶サブシステムに発生した障害を、例えば中間記憶サブシステム１０４cから取得した正制御情報631のハートビートから検出する（１３０１）。そして、副記憶サブシステム１０４ｂは、中間記憶サブシステム１０４ｃの状態を調べる（１３０２）。

【０１３３】

中間記憶サブシステム１０４ｃに対してアクセスが出来ない場合にはこれ以上の更新要求の取得が不可能であるため、副記憶サブシステム１０４ｂはリモートコピーのペア状態をサスペンドへ変更し（１３０８）、差分ビットマップを作成する（１３０９）。

【０１３４】

この場合ビットマップは障害発生時の時点から生成されることになる。図６の例においては、副記憶サブシステム１０４ｂはＤａｔａ２までは更新要求を取得してリモートコピー処理を終了しているため、Ｄａｔａ２以後の情報に関しビットマップへ差分情報を格納する。即ち副記憶サブシステム１０４ｂは、ビットマ

ップを作成後、Ｄａｔａ２以降に副記憶サブシステム104bで更新されたの更新情報に関する差分情報をビットマップに格納する。

【０１３５】

副記憶サブシステム１０４ｂにてデータが更新され、差分情報が作成される場合としては、例えばメンテナンス等のために正記憶サブシステムをサスペンド状態にし、副ホスト１０２ｂが副記憶サブシステムを用いて業務を実施する場合や、本実施例の様に正記憶サブシステムにおいて障害が発生した後に、副ホスト１０２ｂへ業務を移行し、副ホストが副記憶サブシステムに対して入出力要求を発行する場合がある。この際副記憶サブシステム１０４ｂの情報はＤａｔａ２までは確定されており、副ホスト１０２ｂからのライトはＤａｔａ２以降の更新情報として差分情報が保持される。

【０１３６】

図１１の説明で述べた様に、正記憶サブシステムにおいても副記憶サブシステムサスペンド時には、サスペンド後の更新情報について差分情報を保持している。このようにしてサスペンドの後はサスペンド時刻を起点に正記憶サブシステム１０４ａ、副記憶サブシステム１０４ｂの双方にて差分情報が保持されるので、図１４で示すように障害回復後の際同期処理の際には、この差分情報を用いて再同期処理を実行することができる。

【０１３７】

次にステップ１３０２において、中間記憶サブシステム１０４ｃがアクセス可能であると判断された場合に関し説明する。まず副記憶サブシステム104bは、中間記憶サブシステム１０４ｃの制御情報領域１１３をチェックし、副記憶サブシステムが取得していない要求が存在するか調べる（１３０３）。中間記憶サブシステムに要求が存在する場合には、副記憶サブシステム104ｂはキュー領域114から更新要求を取得し、取得した更新要求に従って、副記憶サブシステム１０４ｂのボリュームにデータを反映する（１３０４）。そして副記憶サブシステム１０３ｂは、反映した更新要求だけ制御情報１０７ｂの先頭位置をインクリメントし（１３０５）、再び中間記憶サブシステム104ｃの制御情報領域113を参照することにより、次の要求の有無をチェックする（１３０６）。

【０１３８】

次の更新要求の有無は、（１）副記憶サブシステムが更新要求のヘッダ部分のみを取得して解析した後、更新要求の本体であるデータ部分及び末尾情報を取得する方法と、（２）更新要求から一定サイズ分のデータを副記憶サブシステム１０４ｂが取得し、副記憶サブシステム内で解析する方法とが考えられる。本実施例では（１）を用いて説明するが、本発明ではどちらでも実現可能である。

【０１３９】

副記憶サブシステムはまず、キュー要求のヘッダ部分を取得して、リモートコピーのグループＩＤ、要求ＩＤ、及びタイムスタンプを調べる。副記憶サブシステムはグループＩＤ、キューＩＤを用いて論理的に矛盾がないかを解析し、要求ＩＤを用いて直前の更新要求との連番であるかをチェックし、タイムスタンプを調べて当該タイムスタンプが直前の更新要求のタイムスタンプよりも大きい値かをチェックする。副記憶サブシステムは、調査の結果論理的不整合がなければ更新要求が存在すると判断し、論理的不整合が発生する場合には更新要求が存在しないものと判断する。

【０１４０】

更新要求が存在する場合には、副記憶サブシステムは可能であれば更新要求全体を中間記憶サブシステム１０４ｃから取得し、各ブロックの検査と末尾情報の検査をした後、検査結果が正常な場合に副記憶サブシステム１０４ｂのボリュームにデータを反映する（１３０７）。正記憶サブシステム１０４ａが中間記憶サブシステム１０４ｃへ更新要求を送信している最中に障害が発生した場合には、最後に書き込まれた更新要求に関しては書き込みの途中となってしまう可能性がある。このため、各ブロックの検査や末尾情報の検査により、更新要求の一貫性を確保する。

【０１４１】

その後、副記憶サブシステムは、再度アドレスをインクリメントし（１３０５）、中間記憶サブシステム１０４ｃに次の更新要求が存在するかどうかを再度チェックする（１３０６）。副記憶サブシステムは中間記憶サブシステムに更新要求が存在しなくなるまで、ステップ１３０５からステップ１３０７までの処理を

繰り返す。

【０１４２】

以上の処理を図６を用いて説明する。まず副記憶サブシステム104ｂは制御情報１１３を参照て、キュー領域114からＤａｔａ３，Ｄａｔａ４をステップ１３０４にて中間記憶サブシステムからまとめて取得する。その後ステップ１３０５から１３０７においてＤａｔａ５及びＤａｔａ６各々に関し、更新要求ごとに各々ヘッダ、末尾情報等の解析を行い、論理的不整合が発見されなければData5 及びData6を取得する。そして、副記憶サブシステムは、Ｄａｔａ６の次の更新要求を取得しようとする際に、ヘッダ、データまたは末尾情報のいずれかで情報の不整合を発見し、更新要求の取得を終了する。

【０１４３】

図１４は正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂ間の再同期処理の一例を示す図である。

【０１４４】

まず管理者は管理コンソール１１７にて各装置の状態や、装置間の接続状態を取得し再同期が可能な状態かを調べる（1401）。データ処理システム内の経路の一部や記憶サブシステム１０４が使用不可能となっている等の理由で、再同期が不可能な状況の場合は（１４０２）、エラー処理が実行される（１４１４）。

【０１４５】

管理者はまた、管理コンソールを用いて、データ処理システムの論理的な整合性もチェックする。管理者は、管理コンソールを用いて、全記憶サブシステム１０４について、リモートコピーのグループに属するボリュームが使用可能かを確認し、各グループや各ボリュームに関して障害の検知時刻やビットマップの開始時刻を比較する。管理者は再同期処理に用いられるビットマップの開始時刻を比較することで、データ処理システムの論理的な整合性を確認できる（１４０３）。

【０１４６】

副記憶サブシステム１０４ｂは、例えば図13のステップ1302において中間記憶サブシステムにアクセスできないと判断された場合（1310）の様に、状況により

正記憶サブシステム１０４ａと同等の情報が格納されていない場合がある。従って上述の様に、正記憶サブシステム１０４ａでは副記憶サブシステム１０４ｂによって取得され処理されたことが確認されていない更新要求に関してはビットマップを作成する必要がある。

【０１４７】

例えば、図６の例においては、正記憶サブシステム１０４ａはＤａｔａ１からビットマップの取得を開始するべきである。何らかの論理的不整合においてＤａｔａ６以降の時刻からしかビットマップを取得していない場合には、正記憶サブシステムと副記憶サブシステムとの間で再同期処理をした際にデータの不整合が発生する。この場合Ｄａｔａ３、Ｄａｔａ４、Ｄａｔａ５、及びＤａｔａ６が再同期できないという不整合が発生する。つまり、正記憶サブシステム１０４ａのビットマップ作成時刻（差分情報取得開始時刻）が副記憶サブシステム１０４ｂのビットマップ作成時刻（差分情報取得開始時刻）より新しい場合には（１４０３）、データの不整合が発生するので、エラー処理が実行される（１４１４）。

【０１４８】

データの不整合が生じていないと判断された場合には、再同期処理としてまず、管理コンソール１１７上で取得したシステムの状態を元に、管理者が各リモートコピーのペアを再同期すべく経路等を生成する（１４０４）。この際経路は必ずしも中間記憶サブシステム１０４ｃを介する必要はなく、正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂを直接接続してもよい。

【０１４９】

次に管理者は管理コンソール内に新しい構成情報１０６を作成し、管理コンソール１１７は新構成情報１０６を各記憶サブシステム１０４へ送付する（１４０５）。各記憶サブシステム１０４は送付された新構成情報１０６を元に指示された経路を介して指示された装置への接続を試みる。

【０１５０】

次に再同期の要否が判定される。正記憶サブシステム１０４ａまたは副記憶サブシステム１０４ｂのいずれかに障害が発生しデータが消失した場合には再同期はせずに、初期コピーを行う（１４１３）。

【０１５１】

再同期が必要な場合にはどちらの記憶サブシステム１０４へデータ内容を合わせこむかを判定する（1407）。メンテナンス等で一時的に副記憶サブシステム１０４ｂに業務が移行した場合や、障害の発生により正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ業務を移行した場合には、副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａへ再同期が必要になる。また障害の発生により、副記憶サブシステムでは更新データを反映させることなく、正記憶サブシステム１０４ａにおいて業務を継続した場合には、正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへのデータの再同期が必要になる。即ち、通常は業務を移管したシステム側、若しくは業務を継続して実行していたシステム側にデータ内容を合わせこむ。

【０１５２】

正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ再同期を行う場合には、副記憶サブシステムが有するビットマップを副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａへ送付し（１４０８）、正記憶サブシステム１０４ａにおいて正記憶サブシステム１０４ａのビットマップと副記憶サブシステム１０４ｂのビットマップを統合し再同期用のビットマップを生成する。ビットマップの統合は、具体的には両方のビットマップ上に示されている各ビットの論理和（ＯＲ）を計算することで得られる。

【０１５３】

そして、再同期用のビットマップに従って、副記憶サブシステム内に未だ反映されていない更新要求が正記憶サブシステムから副記憶サブシステムに送信され、最同期処理が実行される（１４０９）。尚、再同期処理開始後は、再同期中においても“Ｃｏｐｙ　ｏｎ　ｗｒｉｔｅ”等の技術を用いることで、記憶サブシステム１０４はホスト１０２から受信するＩＯ処理要求の処理を再開することができる（１４１０）。

【０１５４】

尚、副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａへの再同期処理は、ビットマップの送付方向（１４１１）、更新要求の送信方向（１４１２

）が逆となるだけで、その他は正記憶サブシステムから副記憶サブシステムへの再同期処理と同様の処理によって実行される。

【０１５５】

－第２の実施形態－

図１５にデータ処理システムの別の実施形態を示す。図１５では中間記憶サブシステム１０４ｃ上にはリモートコピーキュー１１４が存在し、構成情報１０６及び制御情報１０７は中間記憶サブシステム１０４ｃには格納されない。構成情報１０６は管理コンソール１１７から正記憶サブシステム及び副記憶サブシステムに送付される。制御情報１０７は中間記憶サブシステム１０４ｃを介さずに正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂとの間で直接やり取りされる。

【０１５６】

第1の実施形態と同様に、正ホスト１０２ａから正記憶サブシステム１０４ａへのデータのライトと同期して、正記憶サブシステム１０４ａから中間記憶サブシステム１０４ｃへ更新要求が送信される。一方制御情報１０７は、ホスト１０２から発行されるライト要求とは非同期に正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ直接送信される。副記憶サブシステム１０４ｂは受信した制御情報１０７ｂを参照して中間記憶サブシステム１０４ｃから更新要求を取得し、更新要求から取得したライトデータを副記憶サブシステムのディスクに格納することによって、リモートコピーを実現する。

【０１５７】

正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂ間で制御情報を直接送受信することで、中間記憶サブシステム１０４ｃを介するために必要となるデータ一貫性チェックや送信遅延を削減することができる。

【０１５８】

－第３の実施形態－

図１６にはデータ処理システムの他の一例として、中間記憶サブシステム１０４ｃを複数有するシステムの例を示す。本実施形態では中間サイト１０１が中間サイト１０１Ａ、中間サイト１０１Ｂと複数存在する。図１６に示すデータ処理

システムにおいては、（１）リモートコピーグループ毎に使用する中間記憶サブシステム１０４ｃを割り当てる、若しくは（２）任意のリモートコピーグループのリモートコピー処理において複数の中間記憶サブシステム１０４ｃを使用する、2種類の実施形態が存在する。

【０１５９】

（１）の実施形態の場合、各リモートコピーグループに着目すると実施の形態１と同様の処理となる。

【０１６０】

（２）の実施形態の場合、正記憶サブシステム１０４ａが更新要求を複数の中間記憶サブシステム１０４ｃに振り分けて送信する処理を実施する。この際各更新要求のＩＤはリモートコピーグループ内、及び中間記憶サブシステム１０４ｃ間に渡り唯一の連続番号である。副記憶サブシステム１０４ｂは複数の中間記憶サブシステム１０４ｃから取得した更新要求を整列した後に、更新要求に従ってデータを自装置が有するディスク装置のボリューム内に反映する。

【０１６１】

本実施形態においては障害時の処理においても特徴がある。本実施形態のように複数の中間記憶サブシステム１０４ｃを有することで、任意の一つの記憶サブシステム１０４が故障した場合でも、データ処理システムの構成情報を再構築すれば、交替パス制御が可能である。交替パス制御処理を行う際には障害が発生した中間記憶サブシステム１０４ｃ上に滞留している更新要求を、正記憶サブシステムから副記憶サブシステムに、別の正常な中間記憶サブシステム１０４ｃを経由して送信しなおす必要がある。再送される更新要求は正記憶サブシステムが過去に障害が発生した中間記憶サブシステムに書き込んだ更新要求であるから、再送される更新要求のIDは、正記憶サブシステム104aが直前に中間記憶サブシステムに送信した更新要求のIDとは連続しない可能性がある。しかし、副記憶サブシステム１０４ｂ側で中間記憶サブシステムから取得した更新要求を要求ID順に整列した後、ボリュームへのデータ格納処理を実行すれば問題は生じない。

【０１６２】

－第４の実施形態－

図１７は交替パスを有するデータ処理システムの一例を示す。図１７に示すシステムには、正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂを結合するリモートコピーリンク１７０１が存在する。リモートコピーリンク１７０１は（１）中間記憶サブシステム１０４ｃの障害時にリモートコピーの交替パスとして、若しくは（２）制御情報通信用パスとして使用可能である。

【０１６３】

（１）中間記憶サブシステム１０４cに障害が発生して中間記憶サブシステム経由ではリモートコピーが実行できなくなった場合には、正記憶サブシステムと副記憶サブシステムは、代わりにリモートコピーリンク１７０１を用いてリモートコピーを実行することができる。リモートコピーリンク１７０１を障害時の交替パスとして使用する場合は、正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂ間は長距離であるため、非同期リモートコピーが用いられる。

【０１６４】

（２）正常時においてはリモートコピーリンク１７０１を用いて制御情報１０７を非同期通信にて送受信することもできる。即ち、正記憶サブシステム１０４ａから副記憶サブシステム１０４ｂへ、リモートコピーリンク１７０１を使用し制御情報１０７を送付することができる。また、更新要求に関するデータ以外のメッセージも、正記憶サブシステムと副記憶サブシステムの間でリモートコピーリンク１７０１を使用して非同期で送受信できる。正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂ間で、中間記憶サブシステムを介することなく直接通信することで，中間記憶サブシステム１０４ｃを介するために必要となる、データ一貫性チェックや送信遅延を削減することが可能である。

【０１６５】

－第５の実施形態－

図１８には副記憶サブシステム１０４ｂを複数有するデータ処理システムの一例を示す。副記憶サブシステム１０４ｂが複数個存在する場合には構成情報１０６や制御情報１０７も、複数の副記憶サブシステム１０４ｂ各々について必要となる。

【０１６６】

具体的には正記憶サブシステム１０４ａから中間記憶サブシステムを介して副記憶サブシステム１０４ｂへ送信される更新要求に関しては一対多の制御が必要になり、中間記憶サブシステムに格納される副制御情報６３２及び副情報６１２は副記憶サブシステム１０４ｂの個数分必要になる。

【０１６７】

副記憶サブシステム１０４ｂは第１の実施形態と同様の制御を行う。正記憶サブシステム１０４ａ側の図１０に示す処理が変更となる。正記憶サブシステム１０４ａは複数の副記憶サブシステム１０４ｂの副制御情報６３２を取得し、処理１００３にて全副記憶サブシステム１０４ｂの副制御情報６３２を比較する。比較の結果、最も更新が遅い（古い）副記憶サブシステム１０４ｂの先頭位置を新先頭位置として新しい制御情報１０７aを生成する。また障害時においても正記憶サブシステムは同様に処理し、処理１１０７のビットマップ生成において差分情報取得開始時刻を更新がもっとも遅い副記憶サブシステム１０４ｂに合わせる。これにより正記憶サブシステム１０４ａと副記憶サブシステム１０４ｂ間で有効な差分情報を保持できるようになる。

【０１６８】

副記憶サブシステム１０４ｂから正記憶サブシステム１０４ａへ送信する更新要求に関しては、一対一対応になるため、第１の実施形態と同様の制御が実行される。

【０１６９】

以上に説明したように、ｎサイト間で実行されるリモートコピー処理におけるコスト、処理の複雑さを削減するために、正記憶サブシステムと副記憶サブシステムとを接続する中間サイトにはI/0処理を実施する通常の中間記憶サブシステムを配置し、リモートコピーに係わる処理は正サイトと正記憶サブシステムと副サイトの副記憶サブシステムが行う。即ち、中間記憶サブシステムはリモートコピー処理時には正記憶サブシステム若しくは副記憶サブシステムからのI/0要求に応じてリード若しくはライトの処理を実行する通常の記憶サブシステムである。

【０１７０】

上記構成において正記憶サブシステムと副記憶サブシステムは中間記憶サブシステムを介してリモートコピーの処理を実行する。正記憶サブシステムと副記憶サブシステムは中間記憶サブシステム内のリモートコピーキュー１１４に、データやメッセージをリードしたりライトしたりすることによって、当該データやメッセージを正記憶サブシステムと副記憶サブシステム間でやり取りする。

【０１７１】

正記憶サブシステムがホストからライト要求と、ライトデータを受信すると、ライトデータ及びライトデータの格納位置等ライト要求に含まれる制御情報を有する更新要求を中間記憶サブシステムに送信し、その後にホストへ応答を返送する(同期リモートコピー)。またリモートコピーキュー１１４の進捗を示す先頭位置等のポインタや更新要求のサイズ等が含まれるキュー制御情報も正記憶サブシステムは中間記憶サブシステムに送信する。制御情報の更新はホストのライト要求とは同期でも非同期でもかまわないが性能向上のため一定間隔毎に非同期に更新するほうが好ましい。尚、中間記憶サブシステムを介すことなく、正記憶サブシステムと副記憶サブシステム間を直接接続して制御情報を直接やり取りしてもかまわない。

【０１７２】

副サイトの副記憶サブシステムは中間記憶サブシステムに書き込まれた正制御情報を読み取り、正制御情報に基づいて中間記憶サブシステム上のキューからライトデータを含む更新要求を取得する。尚副記憶サブシステムは正記憶サブシステムから中間記憶サブシステムへの更新要求の送信とは非同期に、更新要求の取得を行う(非同期リモートコピー)。そして取得した更新要求に含まれる位置情報に基づいて、取得した更新要求に含まれるライトデータを副記憶サブシステム内のディスクに格納する。

【０１７３】

上記の処理において更新要求や制御情報のやりとりは、正記憶サブシステム及び副記憶サブシステムがリード若しくはライト等のコマンド中間記憶サブシステムに発行し、中間記憶サブシステムがこれに応じてリード若しくはライト処理を実行することによって実現される。このため，中間記憶サブシステムはリモート

コピーに係わる機能、プログラム等を必要とせず、低コストでｎサイトリモートコピーを実現することができる。中間サイトの記憶サブシステムとして、より小型の記憶サブシステムサブシステムや、JBOD（Just a Bunch Of Devices)を使用すると一層の低価格化が可能である。

【０１７４】

また、中間記憶サブシステムには更新要求や制御情報を格納するための記憶領域は必要となるが、中間記憶サブシステムが正記憶装置システムや副記憶装置システムが有するデータのコピーを保持し続ける必要はないため、中間記憶サブシステムに必要な記憶容量は正記憶サブシステムや副記憶サブシステムが有する記憶容量より少なくて済む。従って、より少ない記憶容量で、nサイト間でのリモートコピー処理を実行することができる。

【０１７５】

また中間記憶サブシステムは一般に、正記憶サブシステム及び副記憶サブシステムの双方との間でリモートコピー処理を実行する必要があるので、処理負荷が高くなってしまうが、本発明の実施形態においては、中間記憶サブシステムで実行される処理が単純化され、中間記憶サブシステムは正記憶サブシステム若しくは副記憶サブシステムから発行されるI/O要求を処理すれば良いので、中間記憶サブシステムの負荷を低減することができる。更に、中間記憶サブシステムについてはリモートコピーの進捗やステータスのチェック等の処理をする必要はないため、リモートコピーに要する監視・管理処理が軽減される。

【０１７６】

【発明の効果】

本発明によれば、低コストでnサイト間でのリモートコピーを実行することができる。また、本発明によれば、nサイト間でリモートコピーを実行するための処理に要する負荷容量を低減することができる。

【図面の簡単な説明】

【図１】本発明が適用されるデータ処理システムの一例を示す図である。

【図２】記憶サブシステムの構成例を示す図である。

【図３】中間記憶サブシステムが有する記憶領域の一例を示す図である。

【図４】中間記憶サブシステムに格納されるデータの構造の一例を示す図である。

【図５】保護データの一例を示す図である。

【図６】中間記憶サブシステムが有する情報の構成例を示す図である。

【図７】要求のデータ構造の一例を示す図である。

【図８】初期設定の手順の一例を示す図である。

【図９】正記憶サブシステムにおける更新処理の一例を示す図である。

【図１０】制御情報更新処理の一例を示す図である。

【図１１】障害発生時に正記憶サブシステムにおいて実行される処理の一例を示す図である。

【図１２】副記憶サブシステムにおける更新処理の一例を示す図である。

【図１３】障害発生時に副記憶サブシステムにおいて実行される処理の一例を示す図である。

【図１４】再同期処理の一例を示す図である。

【図１５】本発明が適用されるデータ処理システムの他の一例を示す図である。

【図１６】本発明が適用されるデータ処理システムの他の一例を示す図である。

【図１７】本発明が適用されるデータ処理システムの他の一例を示す図である。

【図１８】本発明が適用されるデータ処理システムの他の一例を示す図である。

【符号の説明】

１０２…ホスト

１０４…記憶サブシステム

１０６…構成情報

１０７…制御情報

１０８…更新情報

１１２…構成情報領域

１１３…制御情報領域

１１４…キュー領域

１１７…管理コンソール

２０１…コントローラ

２０２…CPU

２０３…ホストアダプタ

２０４…ネットワークコントローラ

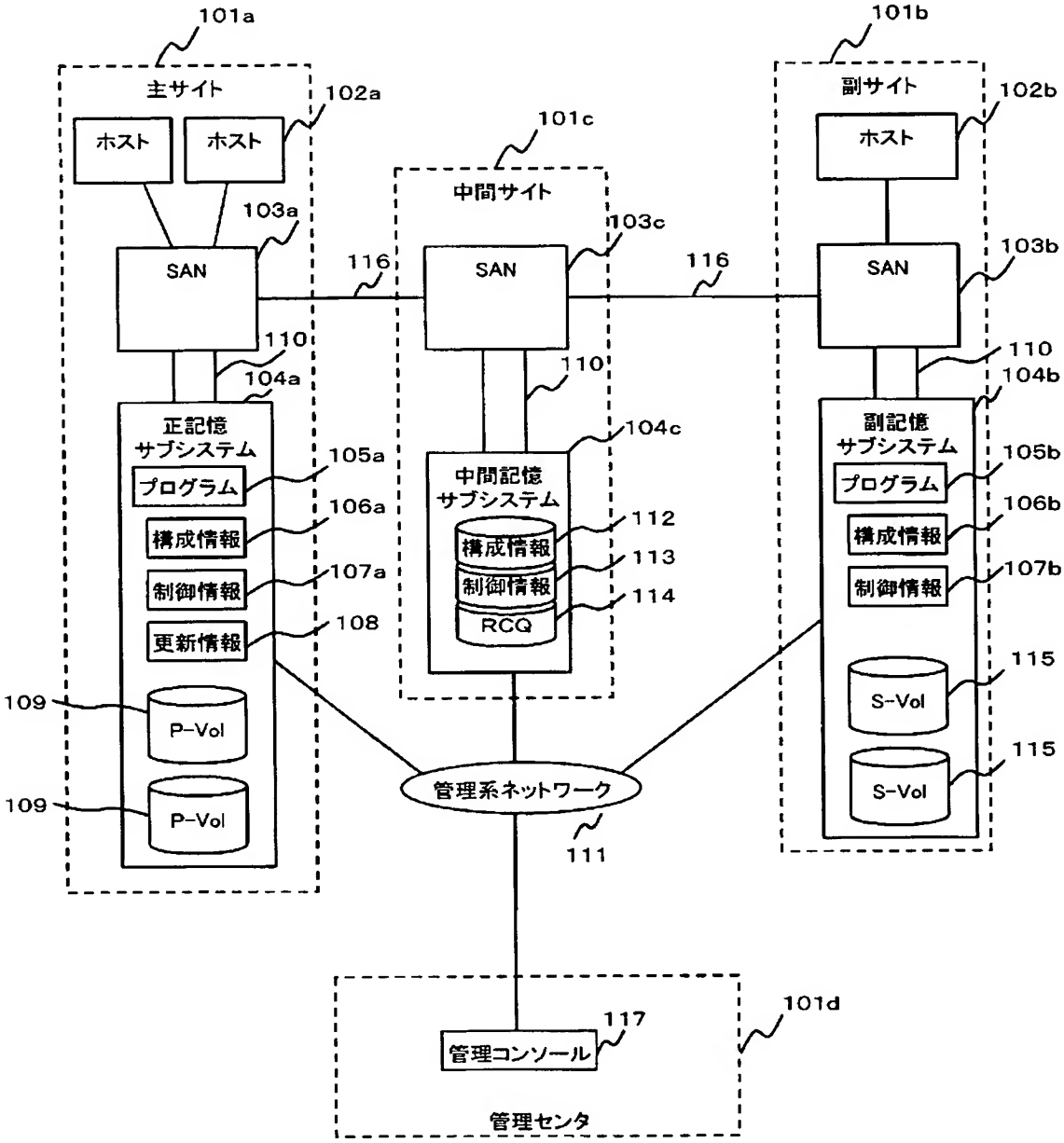
２０５…メモリ

２１０…ディスク装置

【書類名】　図面

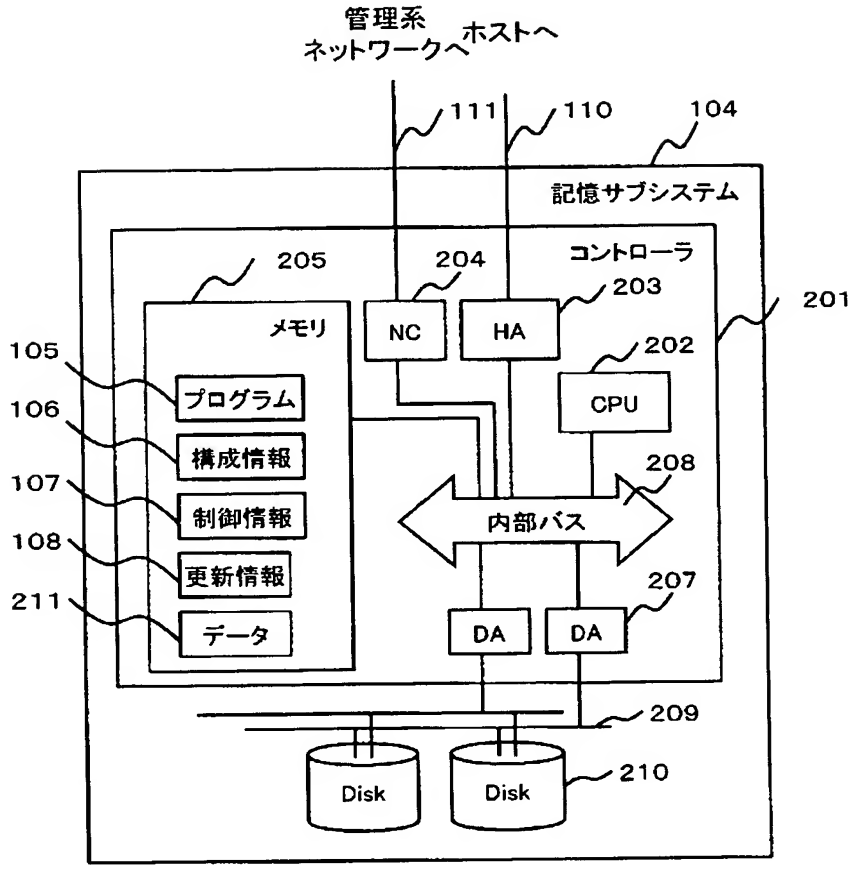
【図１】

図１


101a
主サイト
102a
ホスト
ホスト
103a
SAN
116
101c
中間サイト
103c
SAN
116
101b
副サイト
102b
ホスト
103b
SAN
110
104a
正記憶
サブシステム
105a
プログラム
106a
構成情報
107a
制御情報
108
更新情報
109
P-Vol
109
P-Vol
110
104c
中間記憶
サブシステム
112
構成情報
113
制御情報
114
RCQ
110
104b
副記憶
サブシステム
105b
プログラム
106b
構成情報
107b
制御情報
115
S-Vol
115
S-Vol
管理系ネットワーク
111
101d
117
管理コンソール
管理センタ

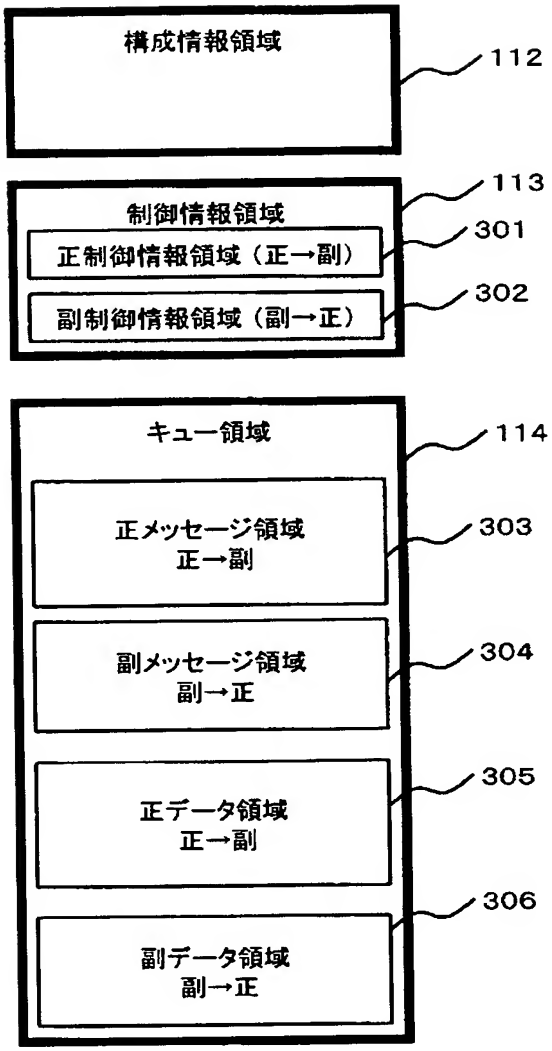
【図２】

図２


管理系
ネットワークへ
ホストへ
111
110
104
記憶サブシステム
205
204
コントローラ
203
201
メモリ
NC
HA
105
202
プログラム
CPU
106
構成情報
208
107
制御情報
内部バス
108
更新情報
207
211
データ
DA
DA
209
Disk
Disk
210

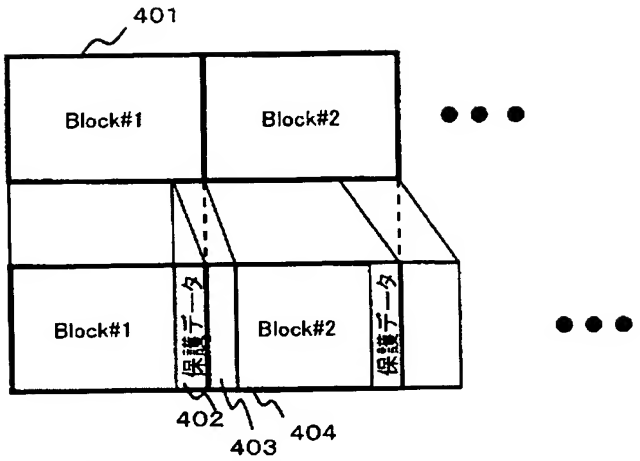
【図３】

図３

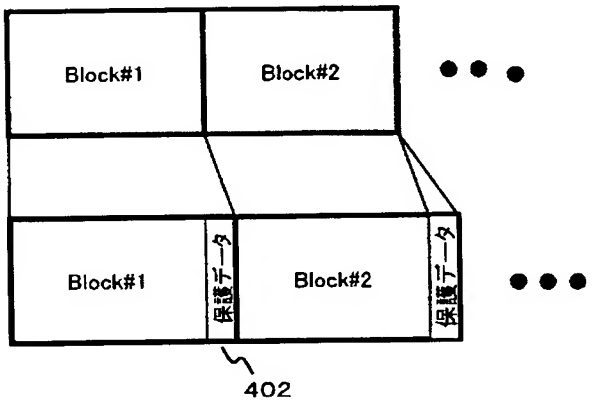

構成情報領域
112
制御情報領域
113
正制御情報領域（正→副）
301
副制御情報領域（副→正）
302
キュー領域
114
正メッセージ領域
正→副
303
副メッセージ領域
副→正
304
正データ領域
正→副
305
副データ領域
副→正
306

【図４】

図４



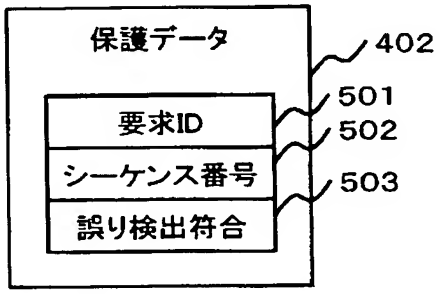
(a) 中間記憶サブシステムにてブロックサイズを変更なし



(b) 中間記憶サブシステムにてブロックサイズを変更

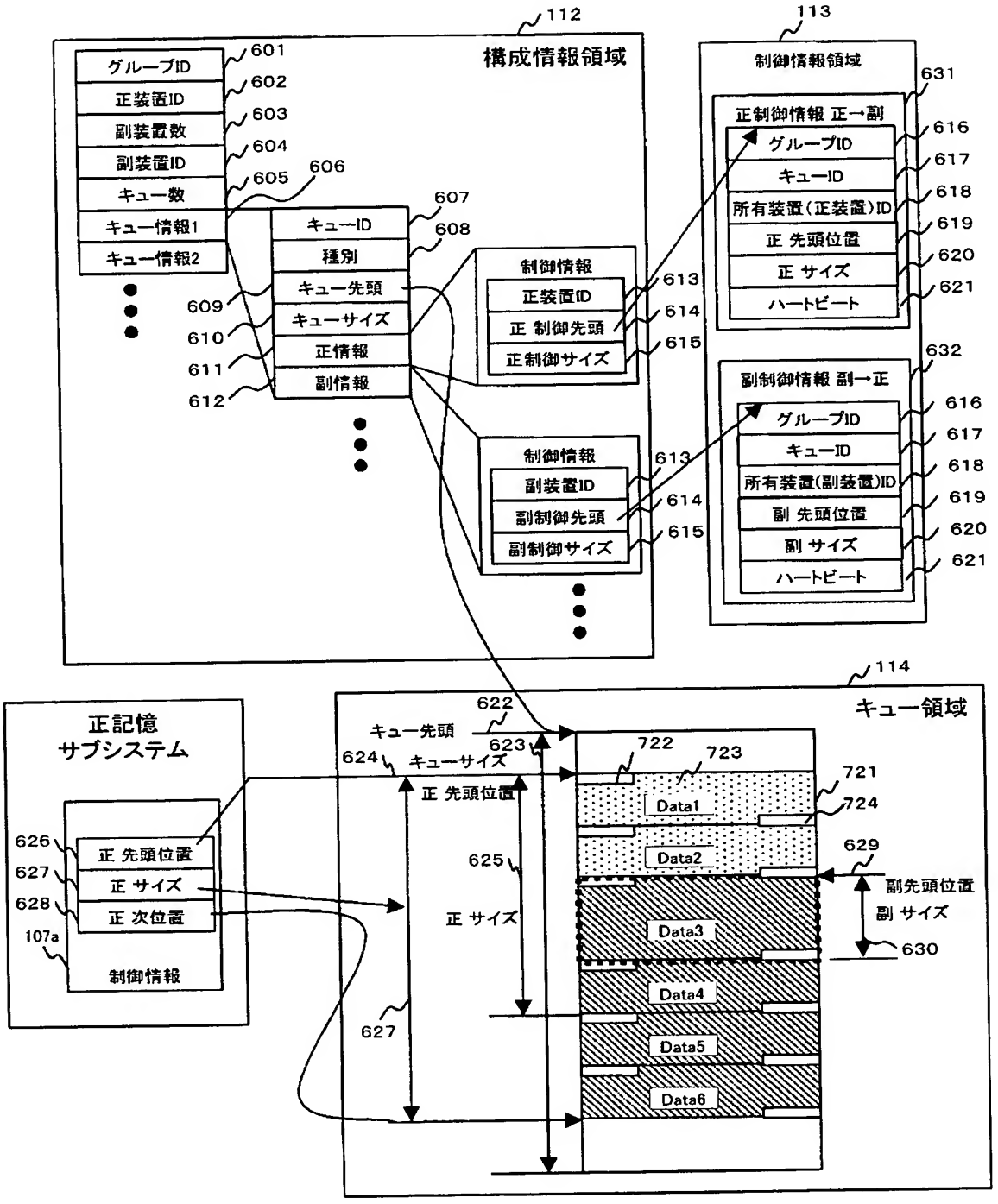
【図５】

図５



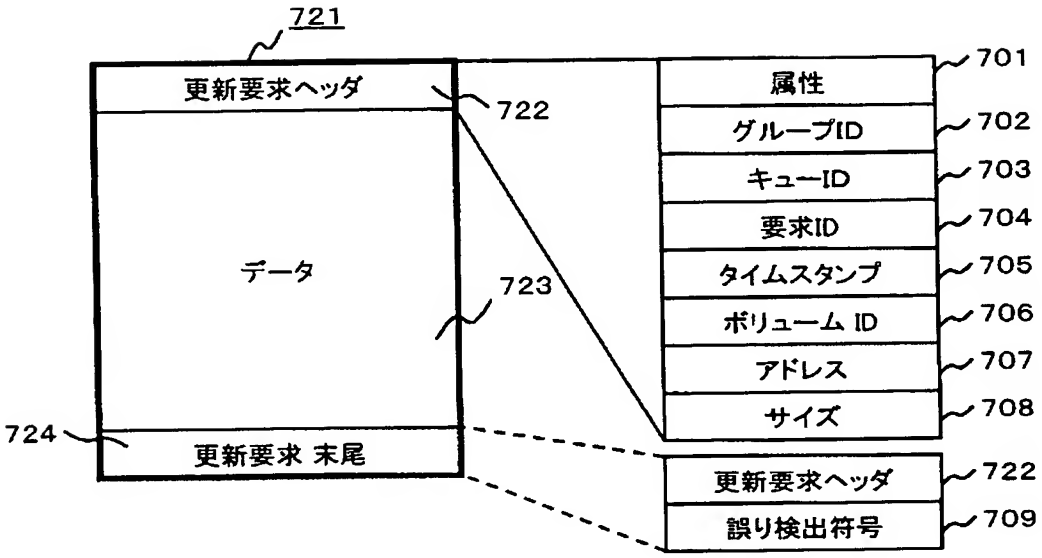
【図６】

図６


112
構成情報領域
601
グループID
602
正装置ID
603
副装置数
604
副装置ID
605
キュー数
606
キュー情報1
キュー情報2
607
キューID
608
種別
609
キュー先頭
610
キューサイズ
611
正情報
612
副情報
制御情報
613
正装置ID
614
正 制御先頭
615
正制御サイズ
制御情報
613
副装置ID
614
副制御先頭
615
副制御サイズ
113
制御情報領域
631
正制御情報 正→副
616
グループID
617
キューID
618
所有装置（正装置）ID
619
正 先頭位置
620
正 サイズ
621
ハートビート
632
副制御情報 副→正
616
グループID
617
キューID
618
所有装置(副装置)ID
619
副 先頭位置
620
副 サイズ
621
ハートビート
114
キュー領域
正記憶
サブシステム
626
正 先頭位置
627
正 サイズ
628
正 次位置
107a
制御情報
622
キュー先頭
623
624
キューサイズ
正 先頭位置
625
正 サイズ
627
722
723
721
724
Data1
Data2
629
副先頭位置
副 サイズ
630
Data3
Data4
Data5
Data6

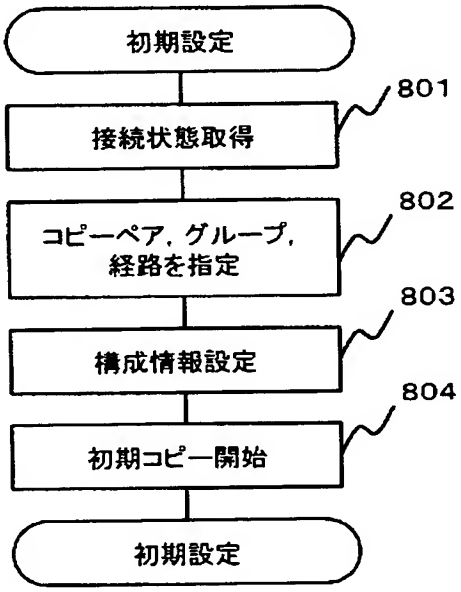
【図７】

図７


721
更新要求ヘッダ
722
データ
723
724
更新要求 末尾
属性
701
グループID
702
キューID
703
要求ID
704
タイムスタンプ
705
ボリューム ID
706
アドレス
707
サイズ
708
更新要求ヘッダ
722
誤り検出符号
709

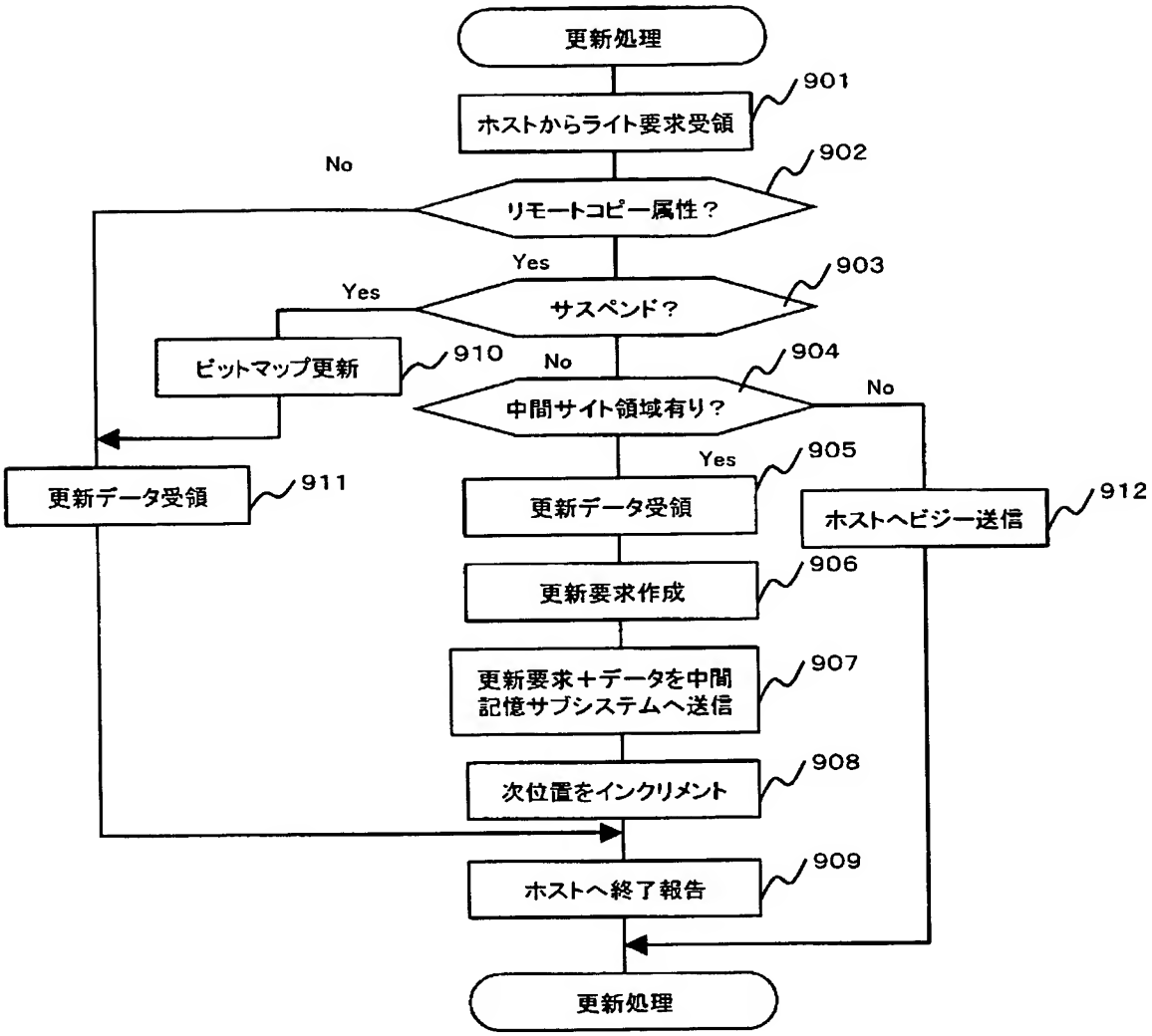
【図８】

図８


初期設定
接続状態取得
801
コピーペア，グループ，
経路を指定
802
構成情報設定
803
初期コピー開始
804
初期設定

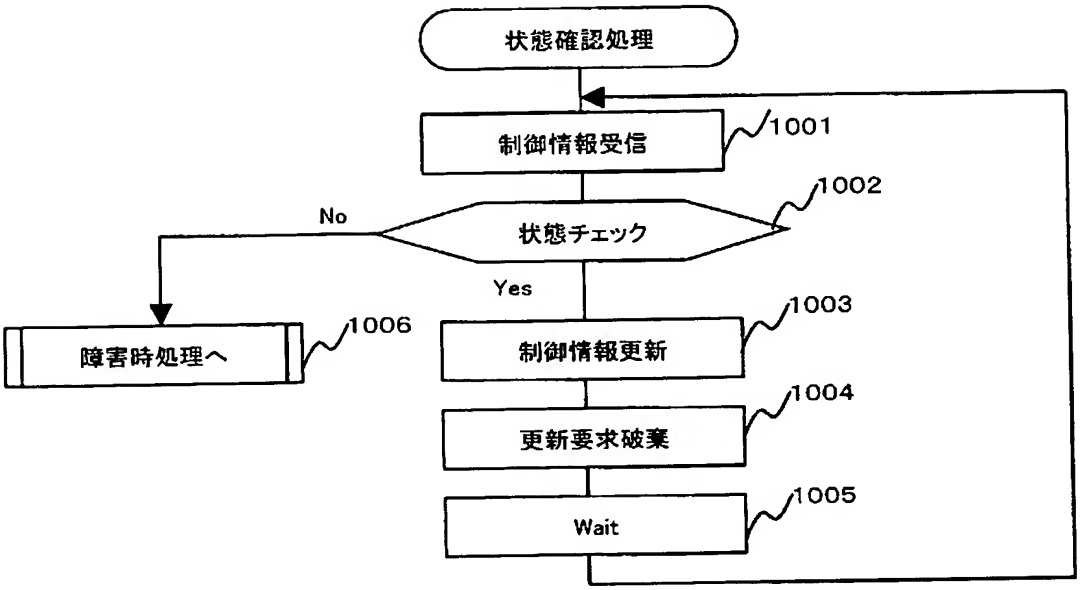
【図９】

図９


更新処理
901
ホストからライト要求受領
902
No
リモートコピー属性？
Yes
903
Yes
サスペンド？
910
ビットマップ更新
No
904
中間サイト領域有り？
No
911
更新データ受領
Yes
905
更新データ受領
912
ホストへビジー送信
906
更新要求作成
907
更新要求＋データを中間記憶サブシステムへ送信
908
次位置をインクリメント
909
ホストへ終了報告
更新処理

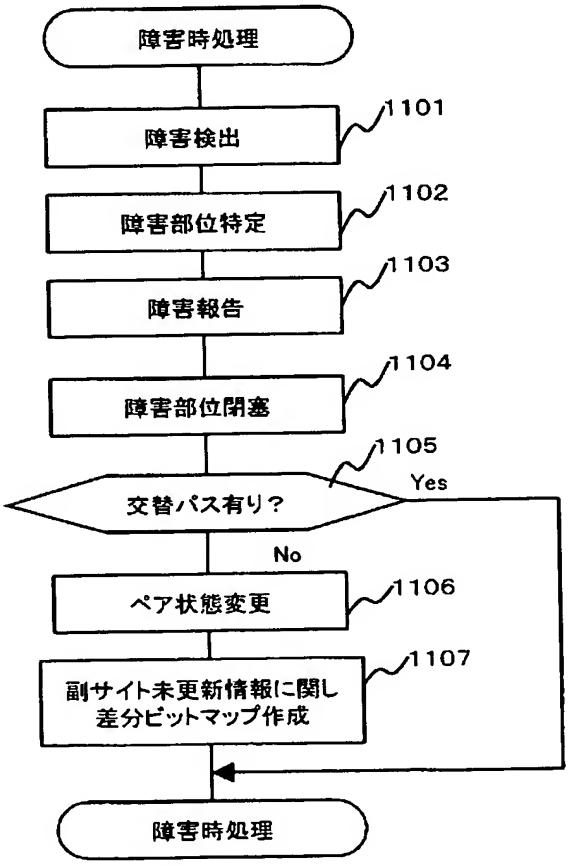
【図１０】

図１０



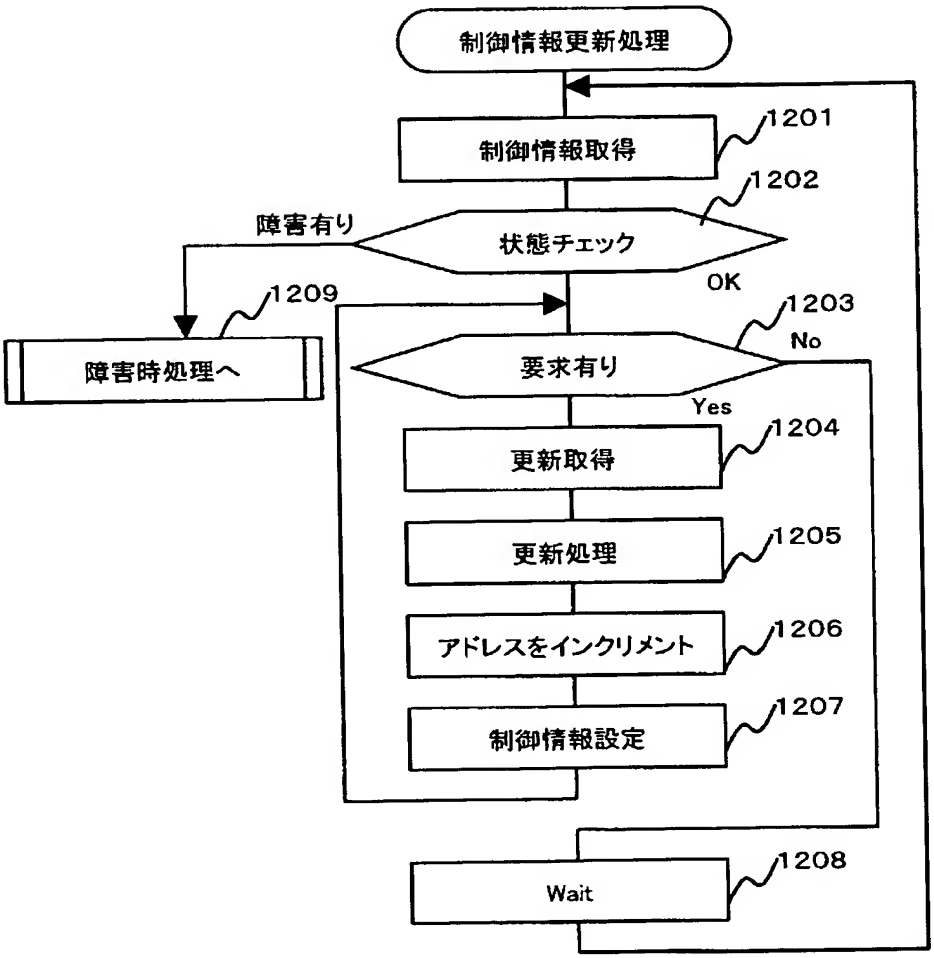
【図１１】

図１１



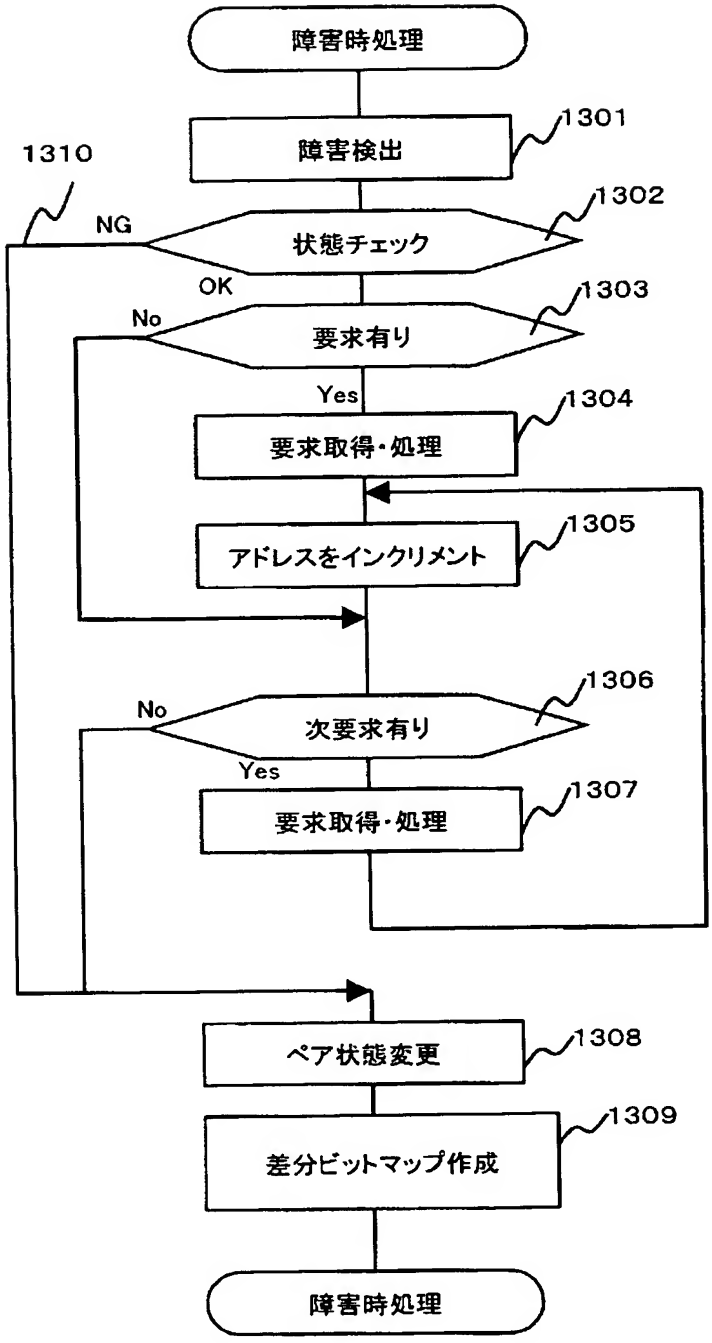
【図１２】

図１２


制御情報更新処理
制御情報取得
1201
状態チェック
1202
障害有り
OK
障害時処理へ
1209
要求有り
1203
No
Yes
更新取得
1204
更新処理
1205
アドレスをインクリメント
1206
制御情報設定
1207
Wait
1208

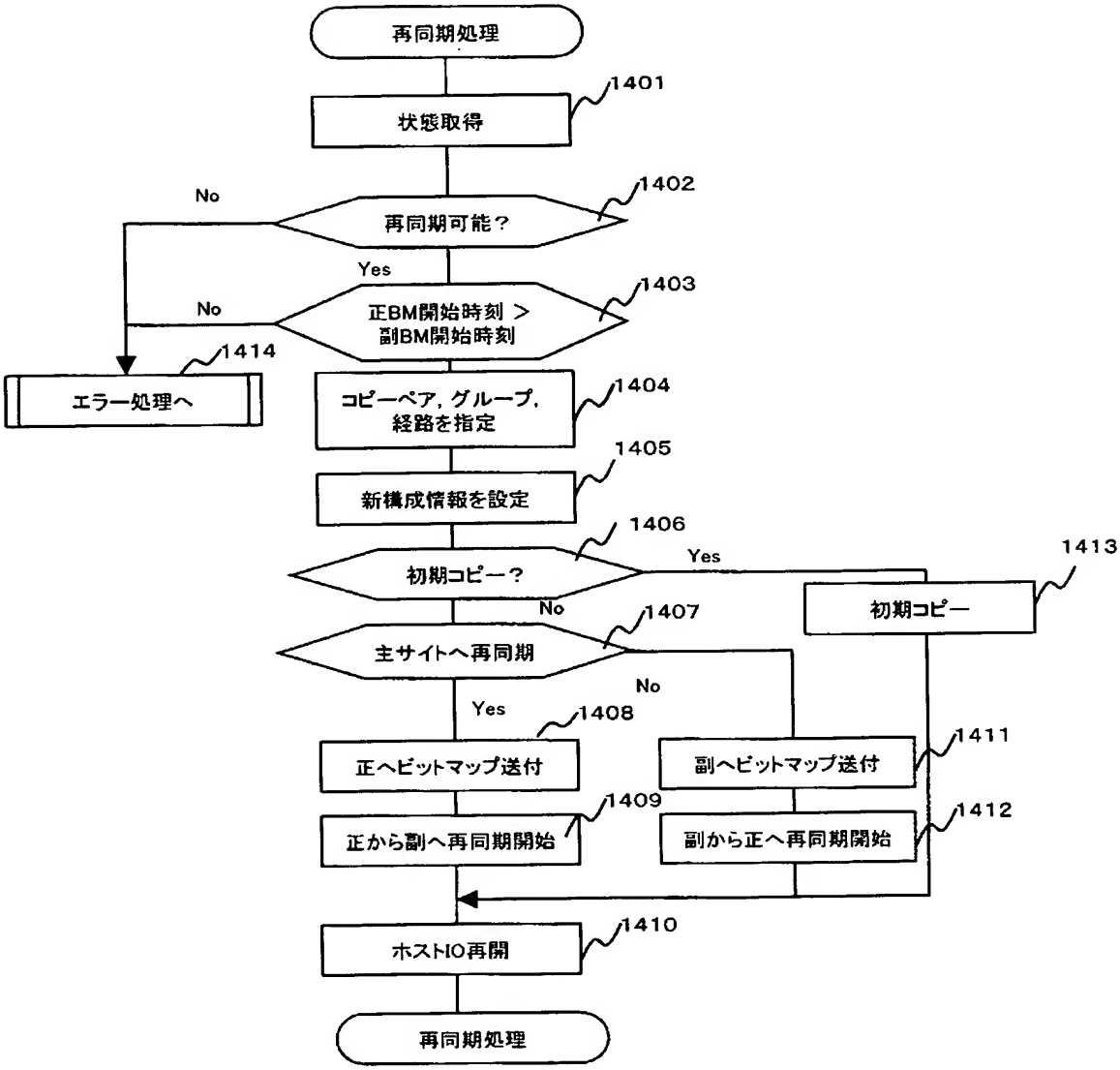
【図１３】

図１３


障害時処理
障害検出
1301
1310
NG
状態チェック
1302
OK
No
要求有り
1303
Yes
要求取得・処理
1304
アドレスをインクリメント
1305
No
次要求有り
1306
Yes
要求取得・処理
1307
ペア状態変更
1308
差分ビットマップ作成
1309
障害時処理

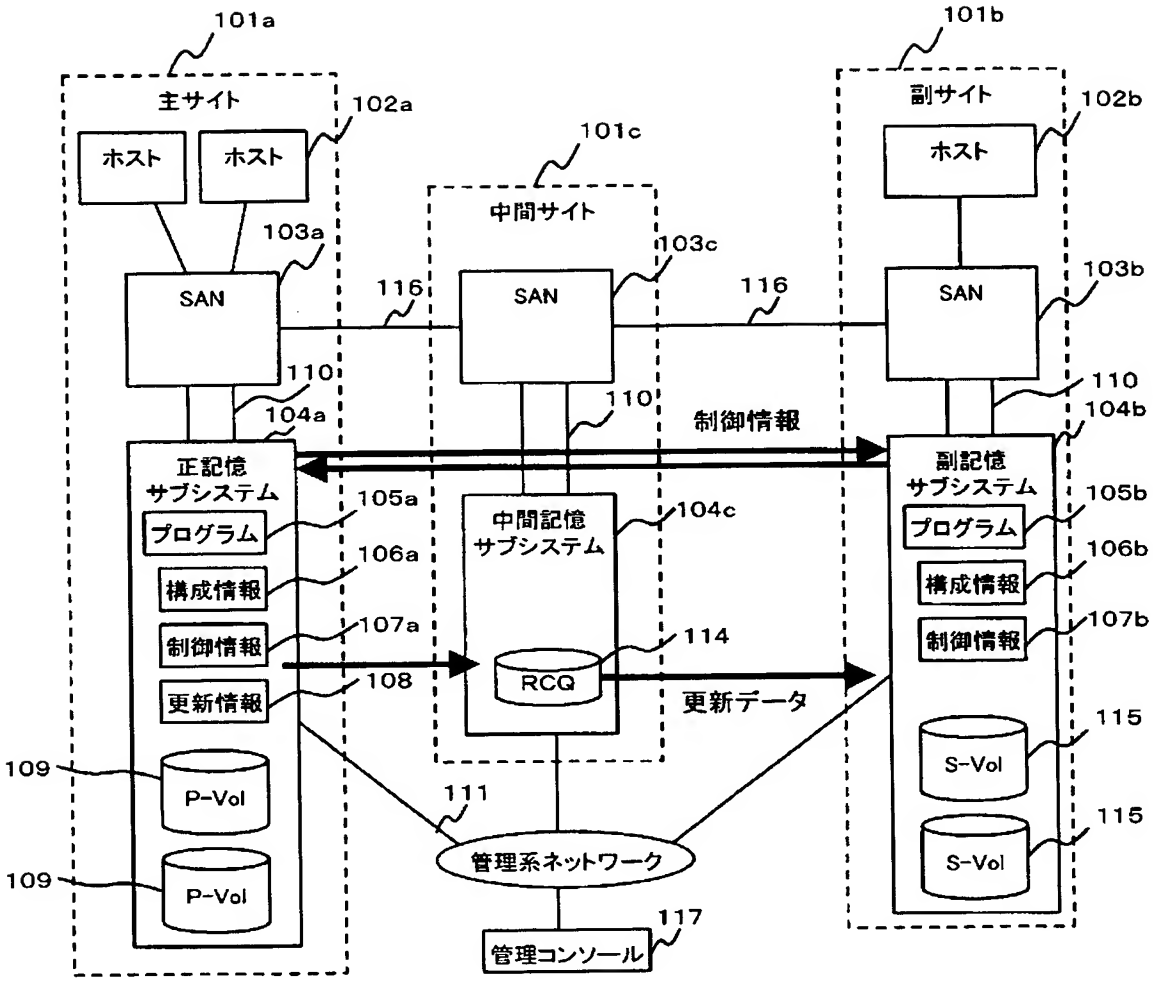
【図１４】

図１４


再同期処理
1401
状態取得
1402
No
再同期可能？
Yes
1403
No
正BM開始時刻 ＞
副BM開始時刻
1414
エラー処理へ
1404
コピーペア，グループ，
経路を指定
1405
新構成情報を設定
1406
初期コピー？
Yes
1413
初期コピー
No
1407
主サイトへ再同期
No
Yes
1408
正へビットマップ送付
1411
副へビットマップ送付
1409
正から副へ再同期開始
1412
副から正へ再同期開始
1410
ホストIO再開
再同期処理

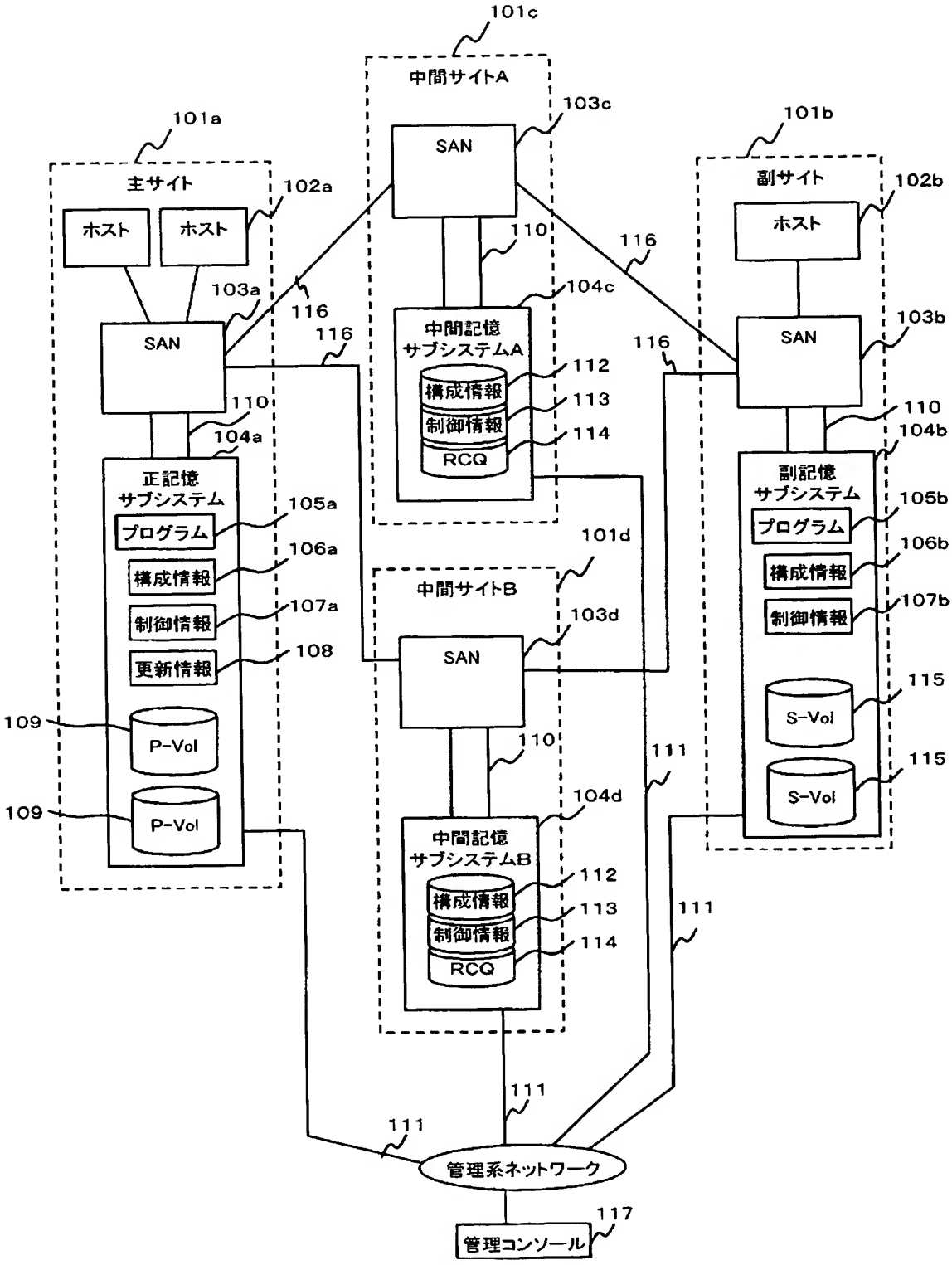
【図１５】

図１５


101a
主サイト
102a
ホスト
ホスト
103a
SAN
110
104a
正記憶
サブシステム
105a
プログラム
106a
構成情報
107a
制御情報
108
更新情報
109
P-Vol
109
P-Vol
101c
中間サイト
103c
SAN
116
116
110
104c
中間記憶
サブシステム
RCQ
114
111
管理系ネットワーク
117
管理コンソール
制御情報
更新データ
101b
副サイト
102b
ホスト
103b
SAN
110
104b
副記憶
サブシステム
105b
プログラム
106b
構成情報
107b
制御情報
115
S-Vol
115
S-Vol

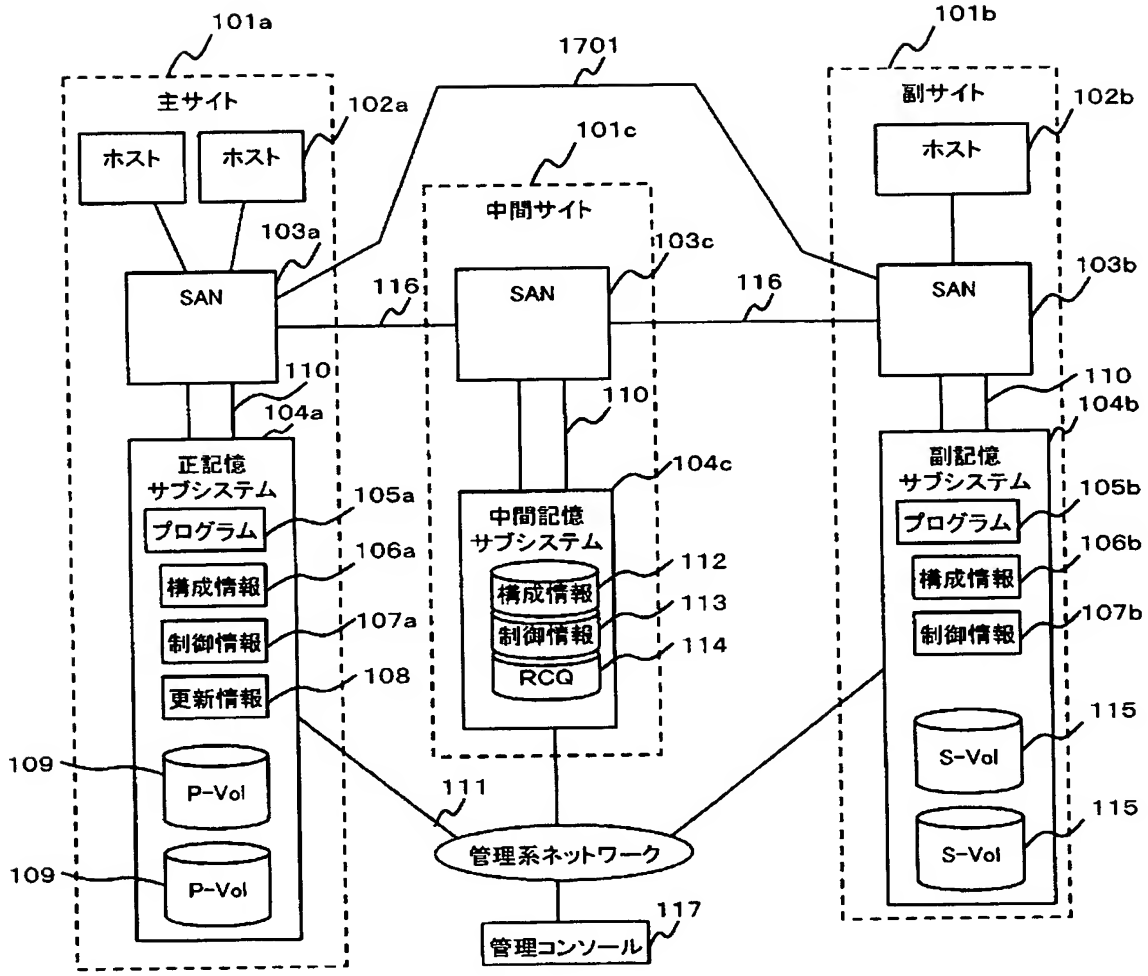
【図１６】

図１６


101c
中間サイトA
103c
SAN
101a
主サイト
102a
ホスト
ホスト
103a
SAN
110
104a
正記憶
サブシステム
105a
プログラム
106a
構成情報
107a
制御情報
108
更新情報
109
P-Vol
109
P-Vol
116
116
110
104c
中間記憶
サブシステムA
112
構成情報
113
制御情報
114
RCQ
116
101b
副サイト
102b
ホスト
103b
SAN
110
104b
副記憶
サブシステム
105b
プログラム
106b
構成情報
107b
制御情報
115
S-Vol
115
S-Vol
101d
中間サイトB
103d
SAN
110
104d
中間記憶
サブシステムB
112
構成情報
113
制御情報
114
RCQ
111
111
111
111
管理系ネットワーク
117
管理コンソール

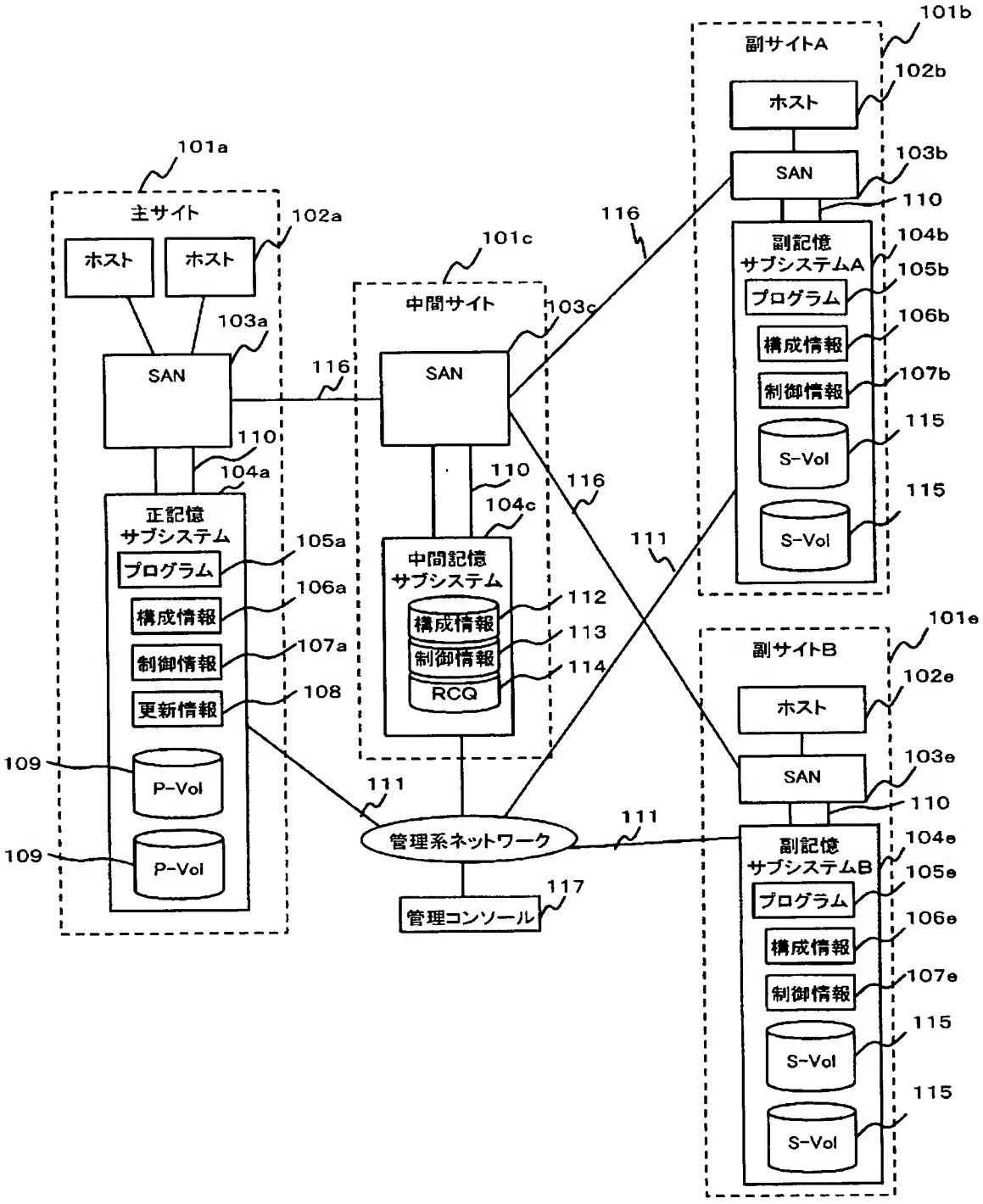
【図１７】

図１７


101a
主サイト
102a
ホスト
ホスト
103a
SAN
110
104a
正記憶
サブシステム
105a
プログラム
106a
構成情報
107a
制御情報
108
更新情報
109
P-Vol
109
P-Vol
1701
101c
中間サイト
103c
SAN
116
116
110
104c
中間記憶
サブシステム
112
構成情報
113
制御情報
114
RCQ
111
管理系ネットワーク
117
管理コンソール
101b
副サイト
102b
ホスト
103b
SAN
110
104b
副記憶
サブシステム
105b
プログラム
106b
構成情報
107b
制御情報
115
S-Vol
115
S-Vol

【図１８】

図１８


101a
主サイト
102a
ホスト
ホスト
103a
SAN
110
104a
正記憶
サブシステム
105a
プログラム
106a
構成情報
107a
制御情報
108
更新情報
109
P-Vol
109
P-Vol
116
101c
中間サイト
103c
SAN
110
104c
中間記憶
サブシステム
112
構成情報
113
制御情報
114
RCQ
111
管理系ネットワーク
117
管理コンソール
116
111
116
101b
副サイトA
102b
ホスト
103b
SAN
110
104b
副記憶
サブシステムA
105b
プログラム
106b
構成情報
107b
制御情報
115
S-Vol
115
S-Vol
111
101e
副サイトB
102e
ホスト
103e
SAN
110
104e
副記憶
サブシステムB
105e
プログラム
106e
構成情報
107e
制御情報
115
S-Vol
115
S-Vol

【書類名】　要約書

【要約】

【課題】

低コストで、処理負荷が小さいnサイトリモートコピーを実現する。

【解決手段】

第１の記憶装置システムと第２の記憶装置システムを第３の記憶装置システムを経由して接続する。リモートコピー処理を実行する際、第１の記憶装置システムは計算機から受信したライト要求に応じ、ライトデータとライトデータが書き込まれる格納位置を示すアドレス情報とを有するジャーナルを第３の記憶装置システムに送信し、第３の記憶装置に書き込みを要求する。第２の記憶装置システムは、第１の記憶装置システムによって発行されるジャーナルの格納位置を含む制御情報を受信し、制御情報に基づいて第３の記憶装置からジャーナルをリードする。そして第２の記憶装置システムは、ジャーナルに含まれるアドレス情報に従って、ジャーナルに含まれるライトデータを、第２の記憶装置システム内のディスクに書き込む。

【選択図】　図１

認定・付加情報

特許出願の番号	特願２００３－２０５６１７
受付番号	５０３０１２８５１３２
書類名	特許願
担当官	第七担当上席　　００９６
作成日	平成１５年　８月　５日

＜認定情報・付加情報＞

【提出日】　平成15年 8月 4日

特願２００３－２０５６１７

出願人履歴情報

識別番号　［０００００５１０８］

１．変更年月日　１９９０年　８月３１日
［変更理由］　新規登録
住　所　東京都千代田区神田駿河台４丁目６番地
氏　名　株式会社日立製作所